

## Dealing with design

**The idea of intelligent design is being promoted in schools and universities in the United States and Europe. Rather than ignoring it, scientists need to understand its appeal and help students recognize the alternatives.**

Scientists tend to tune out when they hear the words 'intelligent design'. The concept, which endeavours to show God's hand shaping the course of evolution, is being promoted in parts of Europe and, more significantly, has recently become popular among Christian fundamentalists who want religion taught in US secondary schools. To most researchers it sounds like politics rather than science, and like someone else's problem.

Mixing as it does the supernatural with scientific doctrine, the concept is a throwback to the days when natural philosophers pursued pseudoscientific disciplines such as alchemy. But the scientific community should not ignore it. As the article on page 1062 reveals, the concept is gaining popularity on US college campuses.

That is because many of the students taught in introductory biology classes hold religious beliefs that conflict, at least on the face of things, with Darwin's framework. Professors rarely address the conflicts between faith and science in lectures, and students are drawn to intelligent design as a way of reconciling their beliefs with their interest in science. In doing so, they are helping it to gain a small, but firm, foothold on campuses around the country.

This is bad news for researchers. Unlike 'creation science', which uses the Bible as its guide, intelligent design tries to use scientific methods to find evidence of God in nature. This approach makes it less theologically heavy-handed than its predecessor, but it also poses a threat to the very core of scientific reason. Most contemporary researchers believe that it is better to keep science and theology firmly separated. Most theologians would agree: intelligent design is not a part of Catholic doctrine, for example.

So what can scientists do to counter the appeal of intelligent design? The concept's advocates frequently approach researchers with offers of campus-wide 'Darwin versus design' debates. Such events

tend to be well attended, but don't change many minds. Furthermore, ill-prepared scientific lectures can sometimes lack the superficial impact of design advocates' carefully crafted talking points.

Scientists know that natural selection can explain the awe-inspiring complexities of organisms, and should be prepared to explain how. But attacking or dismissing intelligent design is likely to aggravate the rift between science and faith that causes students to become interested in intelligent design in the first place.

Scientists would do better to offer some constructive thoughts of their own. For religious scientists, this may involve taking the time to talk to students about how they personally reconcile their beliefs with their research. Secular researchers should talk to others in order to understand how faiths have come to terms with science. All scientists whose classes are faced with such concerns should familiarize themselves with some basic arguments as to why evolution, cosmology and geology are not competing with religion. When they walk into the lecture hall, they should be prepared to talk about what science can and cannot do, and how it fits in with different religious beliefs.

Some will be troubled by the suggestion that they discuss these issues in the classroom. Indeed, it is not the job of a science teacher to meddle with the way their students are brought up or to attack their core personal beliefs. Rather, the goal should be to point to options other than intelligent design for reconciling science and belief.

Even if they manage to sway just a few students, researchers in the United States can have a disproportionate effect on the national debate over science in the classroom. Students often return to their home communities and become teachers, doctors and engineers. It is as local community leaders that those students will become invaluable allies when more conservative religious groups try to halt the teaching of scientific theories in schools. ■

## New accountability in China

**A Chinese funding agency has a new constitution, supporting better selection. Will it spread?**

A new constitution implemented earlier this month by the National Natural Science Foundation of China (NSFC) shows a penchant for transparent and accountable governance. Those who drafted it say it will establish norms and a code of conduct to regulate the foundation's work in a democratic fashion, to establish management based on the law, and to ensure the effective use of funds. It is aimed, as senior NSFC policy-maker Liu Zuoyi put it, at "utilizing overseas brainpower and encouraging overseas scientists to participate in China's basic research" in a "rigorous way".

To draft the constitution, the NSFC studied about 20 legislative documents worldwide, such as the Australian Research Council Act. According to Liu, the constitution guarantees the fairness of funding opportunities by setting out "standards in selecting experts for peer and panel reviews" and "requirements for the management of programmes and results". Failure in such endeavours has been a long-standing complaint of scientists in China and those thinking of returning to the country.

This is all well and good, but the NSFC already has a reputation for fair funding. Scientists' main complaint with its grants has been that the money is too small to pack a punch. In 2004, it handed out 2.8 billion yuan (US\$340 million), a 25% rise over 2003 but still only 1.3% of China's total research and development spending. This money was spread over some 8,300 projects.

Nevertheless, this is a positive step that could make China seem a more comfortable place for the returning scientists on whom the country rests so much of its hopes. But the NSFC's constitution also points to deep problems in other parts of the Chinese funding system. Its impact depends on how well it can inspire other organizations — in particular the Ministry of Science and Technology and the Chinese Academy of Sciences, which account for the lion's share of research funding and hence hold the corresponding responsibility for China's scientific future — to take steps to improve the fair and effective deployment of their money. Will such organizations take up the challenge and follow the NSFC's lead? ■





**Great divide**  
Ecologists play  
tag with tuna  
across the Atlantic  
p1056

**Taking care**  
Common infections  
cause leukaemia  
in toddlers  
p1058

**Unmasked**  
Global warming is  
culprit behind Japan's  
soaring hay fever  
p1059



**Paper trail**  
Intel pays \$10,000  
for original article  
on Moore's law  
p1061

# Turkish government accused of hijacking boosted science budget

**Tamara Grüner, Munich**

The government of Turkey is wresting control of the country's main research council for political ends. That's the accusation of prominent Turkish scientists who fear that recent appointments and legal changes are attempts to channel a growing science budget towards the government's supporters.

This week, parliament is considering the government's second attempt to increase its control over TÜBİTAK, Turkey's main science funding body. The government made its first attempt in 2003, but the law it forced through was later overturned by the country's highest court.

Prime Minister Recep Tayyip Erdogan recently tripled the council's budget to \$300 million, as part of Turkey's negotiations for membership of the European Union. But it is clear that he would like more control over how the money is spent.

TÜBİTAK, set up in 1963 as an independent organization, has an executive board that elects new members, who are then appointed by the prime minister. The board also elects a president, who must be endorsed by the president of the republic, currently Ahmet Necdet Sezer.

The trouble started in 2003, when Erdogan refused to endorse the appointment of six new TÜBİTAK board members. He also refused to pass on to Sezer the board's recommendation that its president, physicist Namik Kemal Pak, should be appointed for a second term. The right-leaning Erdogan and the more left-wing Sezer clashed over the issue and the government quickly passed a law allowing it to appoint unelected members and to name the board's president.

It then appointed six members and an acting president, engineer Nuket Yetis of Marmara University in Istanbul. The new arrivals were not welcome: four vice-presidents resigned, saying that TÜBİTAK had been "taken under political control". And several scientists complained that the new board members were not sufficiently qualified.

The main opposition party challenged the law in the Supreme Constitutional Court, and won in January last year. But Pak and several colleagues are still involved in legal battles to

IMAGE  
UNAVAILABLE  
FOR COPYRIGHT  
REASONS

Prime Minister Erdogan has poured money into Turkish science, but may have jeopardized its autonomy.

get Erdogan's appointments annulled. Pak accuses the government of sacrificing TÜBİTAK's scientific independence: "It changed the law to stop my return," he says.

Turkey's scientific community has been left in disarray. Several directors of research institutes have resigned or been dismissed, including Naci Görür, the director of the main research facility operated by TÜBİTAK, the Marmara Research Centre. They have been replaced by government appointees.

## Legal confusion

With the rejection of the 2003 law, TÜBİTAK's legal status has become unclear. Some TÜBİTAK-funded researchers have already been excluded from international projects after collaborators were advised by lawyers not to get involved, according to Celal Sengor, a geologist at Istanbul Technical University who currently holds the international chair of the Collège de France. "What has happened to TÜBİTAK is a scandal of unprecedented proportion and an affront to science," he says.

To resolve the body's status, the government needs to change the law. Last week, it proposed a law that would divide control of board appointments equally between the government and independent organizations.

In many countries, including the United

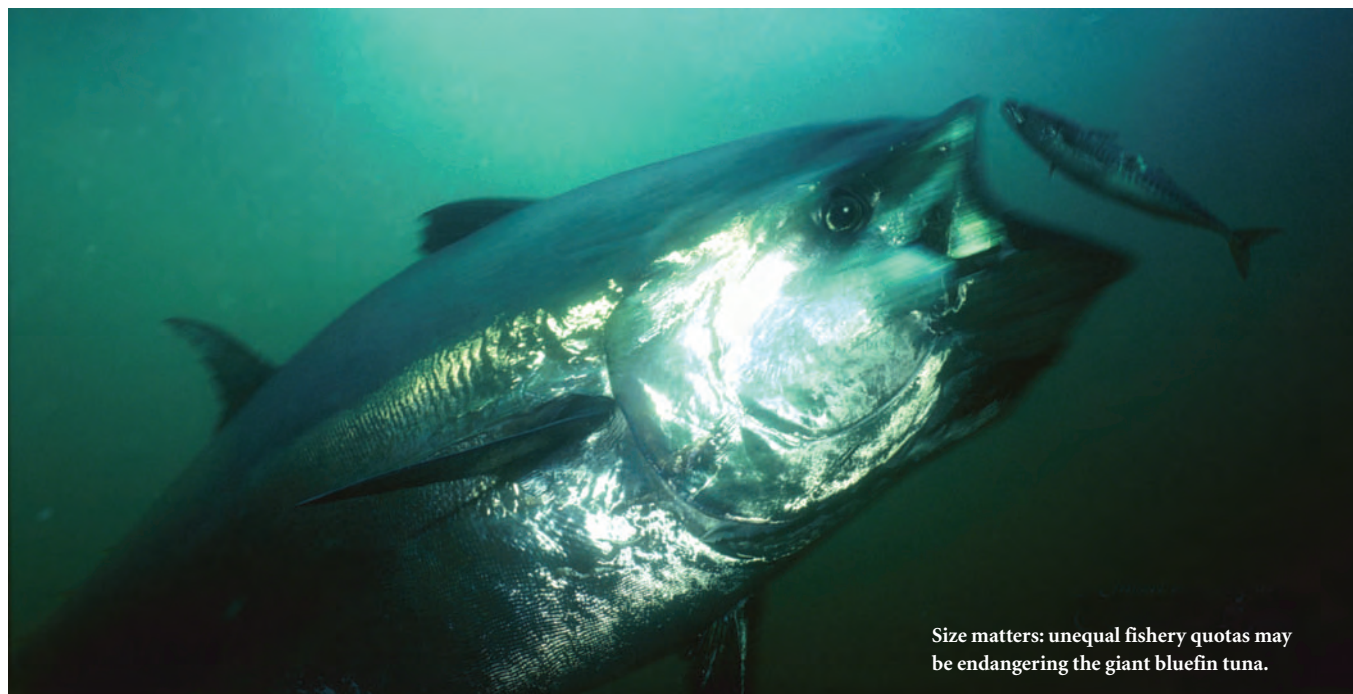
States, governments appoint the officials who run the institutes that distribute public science funding. But decisions about where the money goes are generally supported by a robust system of peer review.

Many scientists in Turkey fear the new law will mean that projects get funding because of political considerations rather than scientific merit. "This would mean the end of independent scientific research," says Sengor.

But Ömer Anlagan, a vice-president of TÜBİTAK and a mechanical engineer at the Middle East Technical University in Ankara, denies that TÜBİTAK's autonomy is under threat: "On the contrary, it will be much better." He argues that the government's moves were necessary to get rid of cronyism. "The old board members always selected the same people." He also denies that the new board members are underqualified. "Some of them are from top universities," he says.

Board member Abdullah Atalar, an electrical engineer from the University of Bilkent, is enthusiastic about TÜBİTAK funding. "We have given 200 young scientists annual grants of \$100,000," he says. "That has never happened before in TÜBİTAK's history." But fellow member Sevkettin Ruacan is uneasy: "It is not certain what the criteria for support were."

A decision on the proposed law is expected in the next fortnight.



Size matters: unequal fishery quotas may be endangering the giant bluefin tuna.

## Satellite tags give fresh angle on tuna quota

**Rex Dalton**

The migration patterns of giant Atlantic bluefin tuna have been unravelled through satellite tagging — and the results suggest that policies for managing the overfished species require an urgent rethink.

Bluefin are huge marine predators that can weigh up to 650 kilograms and are prized for their flesh. They are the most valuable fish in the ocean — in Japan, single fish can command prices of up to US\$100,000. In recent decades, fishing boats have scrambled to catch as many as possible by trap, net, harpoon or long line, and as a result bluefin numbers have fallen by 80% or more since 1970.

Atlantic bluefin (*Thunnus thynnus*) have two spawning grounds — the Gulf of Mexico and the Mediterranean. The gulf is the more heavily fished, so it is given greater protection by the International Commission for the Conservation of Atlantic Tunas (ICCAT), based in Madrid, Spain, which manages tuna stocks in international waters. ICCAT has set bluefin fishing quotas based on an imaginary line down the middle of the Atlantic: fishermen are allowed to catch 32,000 tonnes of bluefin a year to the east of the line, but only 3,000 tonnes to the west.

Scientists and conservationists have long been concerned about this policy, because they suspect that fish cross the dividing line when feeding in the open ocean. But until recently they had no way to prove it.

Now, a tagging study published in *Nature* (see page 1121) provides detailed information about bluefin movements. Getting these data wasn't easy. Over a nine-year period, Barbara Block of Stanford University, Cali-

fornia, and her colleagues caught and tagged nearly 800 bluefin tuna. Global positioning data doesn't work when the fish are on deep dives, so the tags also collected detailed information such as light level and water temperature, which enabled researchers to piece together the course of each fish. Data were beamed back to a satellite, or retrieved after the fish was caught (the team offered a \$1,000 reward for each one returned).

Using the data from the tags, Block and her colleagues confirm that there are two separate populations of bluefin in the Atlantic,

and each returns to the spawning ground from which it came. Crucially, the researchers also find that the two populations merge when foraging in the open ocean, feeding on both sides of ICCAT's line. This means that fish from more fragile western stocks are being caught as part of the larger quotas in the east.

The researchers further showed that longline fishermen hunting the smaller and more abundant yellowfin tuna (*Thunnus albacares*) in the Gulf of Mexico are probably having a devastating effect on spawning

### Bayou fleet brings bluefin blues

Along the Louisiana coast, where bayous provide countless routes to the open water of the Gulf of Mexico, a small fishing fleet is doing great harm to a giant of the sea.

Maybe 50 vessels steam out though the mists from a handful of tiny ports to string their hooks across the waters of the continental shelf in the northern reaches of the gulf. They are seeking the abundant yellowfin tuna that school there nearly all year round. But from April to June the same patch of sea plays host to rare bluefin tuna, which migrate there to spawn. Most bluefin never reach breeding age, which takes around a decade, so catching bluefin at a spawning ground increases the loss of potential offspring. And this makes the endangered species more vulnerable.

Dulac is a typical Louisiana fishing town. Almost all the boats are owned by Vietnamese fishermen, many of them refugees, and they form a cloistered community that is wary of

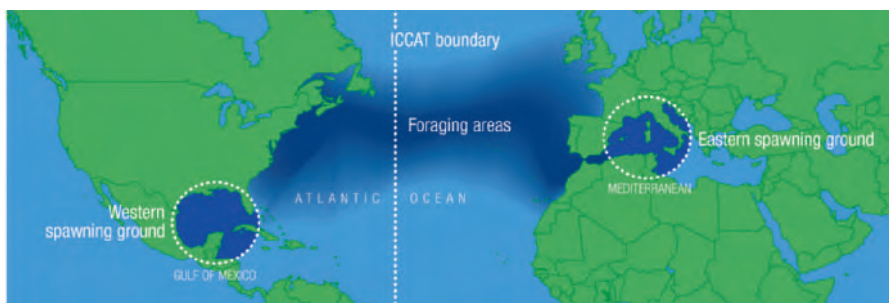
outside inquiry. They go to sea for two weeks at a time, continuously setting their long lines.

Quotas for bluefin are set very low. However, that doesn't really make any difference — even if extra bluefin are thrown back straight away they always die, probably due to the stress of being dragged to the surface in these temperate waters.

Interviews with the fishermen, as well as fish buyers and US officials, indicate that one or two bluefin are killed for every line set. There are even tales of a single line catching ten of the mammoth creatures.

Tung Nguyen (pictured) has just docked his boat, *Peaceful Lady*. He says he has no bluefin aboard, only the 43 yellowfin he caught during his two-week trip. But he admits he does sometimes catch bluefin in the spring, and they weigh anything from 250 to 500 kilograms. He says the fish are spawning — he notices the plump, ripe ovaries of the big females





Fishing for complements: the International Commission for the Conservation of Atlantic Tunas divides bluefin stocks into east and west populations, but there is now evidence that they mix while foraging.

bluefin (see 'Bayou fleet brings bluefin blues'). Although there have been reports of bluefin being caught accidentally in the area, it has been hard for scientists to gauge the scale of the problem.

Block's team combined their location data on the bluefin in the gulf with information from the Maryland-based National Marine Fisheries Service (NMFS) on the movements of fishing boats. They show that the boats are dropping their lines in areas where bluefin gather.

Block hopes the tagging data will be used to set policies that will protect the fish better. "If we don't do something, bluefin stocks will collapse," she says. Block suggests dividing the Atlantic into more than two quota zones, and she wants longline fishing banned in certain parts of the Gulf of Mexico during the bluefin spawning season.

In the past, ICCAT has seemed reluctant to alter its policy, citing the need for more accurate accounts of fish movements. Scientists and conservationists alike hope that this study will fit the bill. "This is a very powerful paper," says John Magnuson, an ecologist at the University of Wisconsin, Madison, who has studied bluefin stocks. "New tagging

methods mean that what was simply a matter of conjecture is becoming visible to us."

"It means we need to rethink bluefin management for the entire Atlantic," adds Paolo Guglielmi, a Rome-based biologist for the conservation group WWF.

The NMFS, which controls fishing practices for much of the bluefin's spawning grounds in the northern Gulf of Mexico, was a part-funder of Block's study and will use the research to revise quotas for bluefin and other migratory Atlantic fish. "We are taking a fresh look," says NMFS economist Rebecca Lent. "We welcome this scientific information."

It is less clear whether ICCAT will change its policies when it sets new bluefin quotas next year. ICCAT officials met in Fukuoka, Japan, from 20 to 23 April to discuss management strategies for tuna and other migratory species. The meeting was partly sparked by Block's preliminary data, which were presented last May in Marseilles, France.

Officials have declined to discuss the outcome of the meeting. But ICCAT's chairman, Masanori Miyahara, cautions that the tagging data should be combined with studies of bluefin ovaries to pin down precisely when and where the fish are spawning. ■



R. DALTON

when he cleans them for market.

Barbara Block of Stanford University, California, and her colleagues have provided the best data so far on the potential impact of the fishery here. They will continue to monitor the fish they have tagged, and as the spawning season begins off Louisiana, researchers around the Atlantic will be watching with more interest than ever. *Peaceful Lady*, and vessels like her, may find that new political forces buffet future trips to sea.

Rex Dalton, Dulac

## Physicists look to crystal device for future of fusion

Mark Peplow, London

Seth Putterman is usually on the side of the sceptics when it comes to tabletop fusion. But now he has created a device that may convince researchers to change their minds about the 'f-word'.

Tabletop fusion has been a touchy subject since Stanley Pons and Martin Fleischmann said in 1989 that they had achieved 'cold fusion' at room temperature. Putterman helped to discredit this claim, as well as more recent reports of 'bubble fusion'.

Now Putterman, a physicist at the University of California, Los Angeles, has turned a tiny crystal into a particle accelerator. When its electric field is focused by a tungsten needle, it fires deuterium ions into a target so fast that the colliding nuclei fuse to create a stream of neutrons.

Puttermann is not claiming to have created a source of virtually unlimited energy, because the reaction isn't self-sustaining. But until now, achieving any kind of fusion in the lab has required bulky accelerators with large electricity supplies. Replacing that with a small crystal is revolutionary. "The amazing thing is that the crystal can be used as an accelerator without plugging it in to a power station," says Putterman.

Puttermann got the idea when he delivered a lecture on sonoluminescence and energy focusing at the Georgia Institute of Technology, Atlanta. Physicist Ahmet Erbil suggested that Putterman should instead consider ferroelectricity.

"Here's someone telling me in front of 100 people that I'm working on the wrong thing," recalls Putterman. But the comment got him started on his fusion reactor. The result is published in this week's *Nature* (see page 1115).

Will he be able to avoid the controversy that has dogged other fusion claims? "My first reaction when I saw the paper was 'oh no, not another tabletop fusion paper'," says Mike Saltmarsh, an acclaimed neutron hunter who was called in to resolve the dispute over bubble fusion. "But they've built a neat little accelerator. I'm pretty sure no one has been able to generate neutrons in this way before."

Puttermann himself isn't worried. "If people think this is a crackpot paper that's just fine," he says. "We're right. Any scientist who says this is too wonderful to believe is welcome to reproduce the experiments." ■

## US experts draw up guidelines for stem-cell research

Erika Check, Washington

US stem-cell research needs a coherent set of rules, according to experts from the National Academy of Sciences. Every institution working with human embryonic stem cells should create a committee to oversee the research, says the panel, which also recommends that some experiments should be banned.

"We think it's very important that everyone who's doing this work is operating in the same spirit, and with the same conditions of transparency and care," says bioethicist Jonathan Moreno of the University of Virginia in Charlottesville, who co-chaired the panel. At present, much research on human embryonic stem cells in the United States is funded from private sources, placing it largely beyond the reach of federal regulations.

In its 26 April report, the panel says that committees should conduct different levels of review for various stem-cell experiments, according to the ethical challenges they pose. "We decided to set up categories to give guidance as to how to think about these things," says panel co-chair Richard Hynes, a cell biologist at the Massachusetts Institute of Technology in Cambridge.

For instance, to study cell development, some scientists want to inject human embryonic stem cells into other species' blastocysts — embryos just a few days old that consist of a hollow ball of cells (see *Nature* 431, 885; 2004). The panel argues that such proposals would require thorough review and that experiments with blastocysts from primates should be banned because of the danger that the 'chimaera' produced could develop to term. "Some things should not be done because they cross ethical lines," says Hynes.

The panel wants at a local level what some other countries have on a national basis. Australia, Britain and Canada have set up national committees to regulate stem-cell research. And in 2002, Singapore's Bioethics Advisory Committee recommended that its government do the same.

In the absence of federal guidelines, it seems the panel's recommendations provide a reasonable blueprint. "I hope they can help us move forward and establish the field," says Leonard Zon of the Children's Hospital Boston, current president of the International Society for Stem Cell Research. ■

IMAGE  
UNAVAILABLE  
FOR COPYRIGHT  
REASONS

Safety in numbers: children who go to day care early in life are less likely to develop leukaemia.

## Link to infection raises hope of preventing child leukaemia

Roxanne Khamis, London

It is finally possible to identify the environmental factors responsible for childhood leukaemia, researchers told a meeting in London last week. Now it's time to think about how to prevent the disease, they said.

Exposure to radiation, chemicals and power lines are not a significant cause, the meeting heard. In fact, most cases are caused by common infections in toddlers.

Leukaemia causes the production of abnormal white blood cells, and accounts for a third of cancers in children. Some genetic predisposition is involved, but for decades scientists have been trying to identify what triggers the disease.

The biggest effort is the United Kingdom Childhood Cancer Study, set up in 1991. It compiled information from more than 10,000 children, including some 1,700 with leukaemia.

Researchers from the project met last week to discuss the results. They were agreed on the role of chemicals and radiation. "Perceived risk factors such as living near sources of electromagnetic fields or natural radiation are not principal causes, if at all, of leukaemia in children," says Mel Greaves of the Institute of Cancer Research in London.

Infections, on the other hand, induce a proliferation of white blood cells in bone marrow as part of the normal immune response. In children genetically predisposed to leukaemia, the researchers think that infection might cause an uncontrolled proliferation of cells, leading to cancer.

Although several studies have hinted that infection could be a cause, it has taken the size and statistical power of the UK pro-

ject to convince researchers that there really is a significant link. To do this, epidemiologist Eve Roman of the University of York and her team analysed data on children's day-care attendance.

Records from East Germany had hinted that infants sent to playgroups from the age of three months were less likely to contract leukaemia. So Roman's team set out to test whether exposure to infections very early in life could somehow train the immune system to protect against the cancer.

They focused on acute lymphoblastic leukaemia, a common form of the disease that usually strikes between the ages of two and five — the time most children start going to playgroups. The team found that children who attended day care during the first three months of life had half the normal risk of developing the disease (C. Gilham *et al. Br. Med. J.* doi:10.1136/bmj.38428.521042.8F; 2005).

Attention is now turning to how to prevent leukaemia. Encouraging parents to send their children to playgroups early in life is one obvious option. Identifying the infections responsible also raises the possibility of developing protective vaccines. Greaves notes that US and Finnish studies have suggested that the Hib vaccine against meningitis also helps to protect against leukaemia.

Charles Stiller, a cancer epidemiologist at the University of Oxford, UK, is excited about the shift but cautions against getting too carried away. "It's difficult to know what proportion of cases is accounted for by infections," he says. "And there is more to be done on defining the mechanisms by which this might work." ■

# Climate change blamed for rise in hay fever

**Rachael Williams, Tokyo**

Spare a thought for Japan's myriad hay-fever sufferers as they endure the highest pollen levels on record this spring. Global warming seems at least partly to blame and most experts agree that the worst is yet to come.

Hay fever in Japan is more punishing than that triggered by weed pollen, which occurs in much of the rest of the world. It is caused by an allergic reaction to cedar and cypress pollen. Severe symptoms and the spread of pollen over wide distances is forcing thousands of people, even in urban areas, to don protective masks and glasses.

There are about 7 million hectares of cypress and cedar plantations in Japan. In Tokyo alone, about a quarter of the population is suffering from hay fever. And according to government surveys the number is steadily rising.

Kouji Murayama, a researcher at the Japan Meteorological Business Support Center in Tokyo, believes that the culprit is global warming. He points to studies that show a clear link between summer temperatures and the amount of pollen produced the following spring. Such data already provide the basis for pollen forecasts.

Tokyo's average yearly temperature has increased by 3 °C since 1890 and, according to the Japan Meteorological Agency in Tokyo,

IMAGE  
UNAVAILABLE  
FOR COPYRIGHT  
REASONS

Fever pitch: many Japanese need protective masks to help them fend off the effects of tree pollen.

is set to rise by up to 3.5 °C by the end of this century.

Based on this forecast, Murayama predicts that the number of hay-fever sufferers in Japan will rise 40% by 2050. "Global warming

will continue to intensify what is already a serious health problem in Japan," he says.

Many agree with Murayama's findings. "It's common sense," says Atsushi Ueda, an allergy specialist at Kumamoto University. Ueda adds that higher levels of carbon dioxide and diesel-exhaust particles can also worsen the body's response to pollen.

But other scientists argue that economic factors may be to blame. Yoko Fukuda, a researcher at Japan's Forestry Agency, points to the decline in domestic forest industries, which has left the cedar and cypress plantations unmaintained. "Such neglect has allowed the trees to mature to their prime pollen-producing age," he explains.

To comfort the hordes of sufferers, the Forest Tree Breeding Center is working on a plan to replace all the offending trees with pollen-free cypress and cedar. But this could take decades, and political support is still uncertain, says project leader Makoto Takahashi.

Despite the difficulties, scientists are determined to find ways to improve the situation — and to ease their own symptoms. Murayama's research was first motivated by his wife's hay fever. But after conducting studies in the forests, he has experienced the problem first hand. "I can now tell you all about the miseries of pollen allergies," he says. ■

# Corporate culture nets big bucks for university heads

**Emma Marris, Washington**

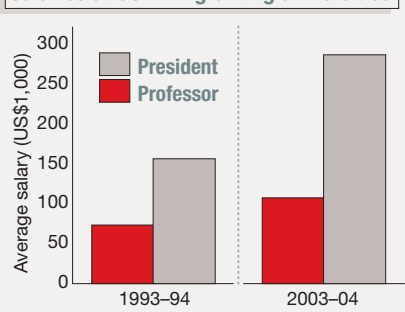
A growing tendency for US universities to embrace private-sector attitudes is undermining the system, say many of the country's professors. Many campus presidents are now paid as if they were corporate chief executives — and some academics say that this is rewarding the wrong type of leadership.

A recent report has revealed that salaries for university presidents have shot up over the past decade, whereas pay packets for professors have barely kept pace with inflation (see graph, right).

John Curtis, director of research at the American Association of University Professors (AAUP) in Washington DC, discovered the extent of the disparity when he compared figures from the AAUP's annual salary survey with presidential salary data from the College and University Professional Association for Human Resources.

The average professor at a PhD-granting university earned about \$100,700 in 2003–04, he says, whereas the average president got \$270,000. The difference is especially pronounced at private institutions, he adds.

Salaries at US PhD-granting universities



Johns Hopkins University's president, William Brody, was the most highly paid president in 2003, according to a survey in *The Chronicle of Higher Education*, earning just over \$590,000.

Curtis sees the growing gap as a "move towards the more corporate style of management", with universities paying high salaries for ruthless cost-cutting leadership. This is the wrong model, he argues, because universities are not run to make shareholders rich, but to educate students and benefit society.

Raymond Pierotti, a biologist at the

University of Kansas in Lawrence, has been following the salary trends for several years. He says that a dip in federal and state funding is forcing universities to act more like companies. He worries that the high salaries will attract "people who are willing to sell out what a university is supposed to stand for because they are more motivated by money".

But others say that the wages are fair compensation for an increasingly tough job. "The demands of the job are expanding," says Melanie Corrigan, a policy analyst at the American Council on Education in Washington DC. "A president is a cross between a chief executive of a large corporation and a small-city mayor."

Pierotti says that the trend won't be reversed unless there is an influx of cash from government sources. But this may be far off — state and federal budgets are under pressure and are likely to remain tight for the foreseeable future. Meanwhile, Pierotti warns that the bottom line is beginning to affect professors as well. "Faculty members are assessed not only on the quality of their teaching or even their research, but on how fundable their research is," he says. ■



## Boycott of Israeli universities sparks academic backlash

**London** British academics have voted to boycott two Israeli universities.

The Association of University Teachers will boycott Bar-Ilan University because of its links with a college in Ariel, an Israeli settlement on the West Bank. Delegates at the association's annual council meeting in Eastbourne last week also voted to sever links with Haifa University following claims that administrators there have tried to censor academics who are critical of Israeli government policies.

The boycott motion was criticized by the academic groups that campaigned against a previous attempt at a broader boycott of Israeli institutes. Such sanctions curb academic freedom and thwart attempts to secure a lasting peace in the region, says Scholars for Peace in the Middle East, a largely US-based academic network.

## United front boosts Asian stem-cell research

**Tokyo** Scientists behind the first Asia-Pacific stem-cell network said last week they are confident of avoiding the ethical and legal problems that hinder such work in the West.

European human embryonic stem-cell scientists are hampered by wide variations in religious and legal approaches across the continent (see *Nature* 434, 809; 2005), and their US colleagues face tough restrictions over how they can use federal government money. But stem-cell scientists from seven Asia-Pacific nations, who met last week in Japan and agreed to collaborate, predicted that their region's more uniform and permissive attitudes towards their work will ensure they do not face such problems.

The group, tentatively titled the Asia-Pacific Stem Cell Network, will host symposia, encourage collaborations and exchange students. All the member countries — Japan, China, Hong Kong, Taiwan, Singapore, South Korea and Australia — have leading stem-cell researchers.

## Space debris collisions hit unexpected high

**Munich** Sobering news for those concerned about space debris: researchers learnt last week of a crash in January between the discarded upper stages of a US and a Chinese rocket.

This brings the total number of collisions in orbit since 1991 to three, although models had predicted only one such event every 14 years. More data are needed to determine whether this is a statistical fluke, or a sign that the model is underestimating





## Hubble marks birthday with fresh image

**Washington** The US and European space agencies have celebrated the birthday of the Hubble Space Telescope, which is 15 this week, by releasing new images of the Eagle Nebula, and the Whirlpool Galaxy M51.

The pictures of the Eagle Nebula (left), one of the telescope's most famous subjects, capture a tower of cold gas and dust, almost ten light years high, rising from an area where new stars are being born. Ultraviolet light from nearby stars causes oxygen to glow blue and hydrogen red.

In total, astronomers have Hubble to thank for almost 750,000 images, more than 100 gigabytes of data a week, and a much-improved understanding of the farthest recesses of the Universe.

Hubble is due to be succeeded in 2011 by the James Webb Space Telescope, but NASA is currently undecided whether to replace Hubble's batteries when they fail in 2007 or 2008.

collision rates, said delegates at the European Conference on Space Debris, held in Darmstadt, Germany, on 18–20 April.

Researchers added that the increasing danger to the 700 active satellites is not being taken seriously enough. There are around 10,000 objects larger than 10 cm in Earth's orbit and the number is rising. Collisions and explosions caused by excess fuel have generated a further 100,000 smaller fragments.

## Plant-to-pill pharming takes transgenics underground

**Washington** Fears of gene transfer have dogged attempts to create transgenic crops that produce pharmaceuticals. No one wants livestock vaccines turning up in breakfast cereal, for example. But underground agriculture could provide one solution, according to the latest results from an experiment in which corn, or maize, was grown in a disused quarry.

The first crops have been illuminated for about a year by artificial light and are doing well, says Cary Mitchell, a horticulturalist at Purdue University in West Lafayette, Indiana. Mitchell's team presented results on the crop at a US Department of Agriculture (USDA) meeting held last month in Tucson, Arizona.

No drug-producing crops have been tried yet, but the team hopes eventually to see the plants turned into pills in a seamless underground process. The idea could work, says Norman Ellstrand, a gene-flow expert at the University of California, Riverside. "But if they don't field-release it, they don't have to report it to the USDA, which is a little worrying," he adds.

## Briton wins Moore's law paper chase

**Washington** Who said paper was dead? Forty years ago this month, Gordon Moore made his famous prediction in *Electronics* magazine that the number of transistors in integrated circuits would double every year. Not content with an electronic version, computer-chip firm Intel last week paid US\$10,000 for a 19 April 1965 issue to mark the 40th anniversary of Moore's law.

No sooner had the bounty been offered on 11 April than the volume containing the prize issue was filched from Grainger Engineering Library at the University of Illinois, Urbana-Champaign, despite Intel's warning that it would be on the lookout for stolen library copies. Several universities acted quickly to lock up their own copies.

Intel finally purchased the issue from David Clark, a British engineer who had kept old issues of the magazine.



Bargain basement? David Clark sold an old copy of *Electronics* magazine for US\$10,000.





# Who has designs on your students' minds?

The intelligent-design movement is a small but growing force on US university campuses. For some it bridges the gap between science and faith, for others it goes beyond the pale. Geoff Brumfiel meets the movement's vanguard.

For a cold Tuesday night in March, the turnout is surprisingly good. Twenty or so fresh-faced college students are gathered together in a room in the student union at George Mason University in Fairfax, Virginia, the state's largest public university. They are there for the first meeting of Salvador Cordova's Intelligent Design and Evolution Awareness (IDEA) club.

"I have a great deal of respect for the scientific method," Cordova tells his attentive audience as he outlines the case for intelligent design. Broadly speaking, he says, the concept is that a divine hand has shaped the course of evolution. The arguments are familiar ones to

both advocates and opponents of the idea: some biological systems are too complex, periodic explosions in the fossil record too large, and differences between species too great to be explained by natural selection alone. Cordova — who holds three degrees from the university, the most recent one in mathematics — argues that the development of life on Earth would be described better if an intelligent creator is added to the mix.

Most scientists overwhelmingly reject the concept of intelligent design. "To me it doesn't deserve any attention, because it doesn't make any sense," says Bruce Alberts, a microbiologist and president of the National



Salvador Cordova sets out the basic principles of intelligent design at a campus meeting.

Academy of Sciences. "Its proponents say that scientific knowledge is incomplete and that there's no way to bridge the gap except for an intelligent designer, which is sort of saying that science should stop trying to find explanations for things."

But despite researchers' apparent lack of interest, or perhaps because of it, the movement is catching on among students on US university campuses. Much of the interest can be traced to US teenagers, more than three-quarters of whom believe, before they



reach university, that God played some part in the origin of humans (see graphic, right). But others are drawn to the idea out of sheer curiosity.

"Students are in the challenge-authority mode at that time in their life, and I think they're intrigued," says Stephen Meyer, director of the Center for Science and Culture at the Discovery Institute, the nation's largest intelligent-design think-tank in Seattle, Washington. Since the first IDEA club began at the University of California, San Diego, in 1999, more than 20 chapters have opened on college campuses around the country. In addition, a small number of academics have begun to offer courses on intelligent design (see 'Cast out from class', overleaf).

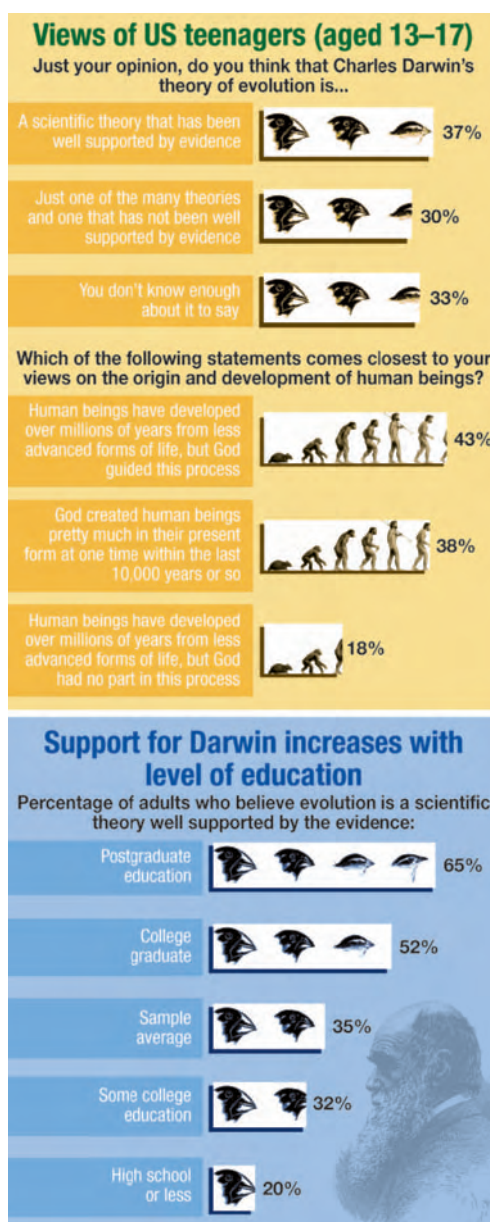
### Relative unease

Darwinists are divided in their response to the idea's growing profile on campus (see 'Natural divisions', overleaf). Many feel that the very presence of intelligent design in universities is legitimizing the movement and eroding the public's perception of science. "Intelligent-design advocates want to split open the public's understanding of science and convince people that you can call on the supernatural for a scientific explanation," warns Barbara Forrest, a philosopher at Southeastern Louisiana University in Hammond and co-author of *Creationism's Trojan Horse: The Wedge of Intelligent Design*.

But others feel that the movement deserves an airing at the university level, even if they oppose its teaching in public schools. "I think that college is a place for experimentation," says Eugenie Scott, director of the National Center for Science Education, a group based in Oakland, California, that promotes the teaching of evolution in public schools. If intelligent design is gaining ground on college campuses, she adds, then scientists are as much to blame as anyone. "I think college professors can do a better job of teaching evolution," she says.

Back in the student union, Cordova is carefully pointing out what intelligent design can, and can't, do. The concept makes no attempts to verify the creation myth or other major biblical events, such as the flood, he says. Nor does it worry about whether Earth is a few thousand years old, as most creationists believe, or four-and-a-half billion years old, the current geological estimate. Intelligent design, Cordova notes, does not even attempt to prove the type of deity involved, it just points to some sort of supernatural intervention. In other words, he says: "Intelligent design doesn't have any theology to it."

It is that distinction that has helped propel the small community of intelligent-design proponents to the forefront of US politics. In 1987, the US Supreme Court struck down a Louisiana law that mandated



teaching 'creation science' in schools because the premise of the research was based on biblical texts. As intelligent design does not draw directly from biblical sources, Christian fundamentalist groups have seized on it as a possible way to force creationism back into the classroom. Last October, a school board in Dover, Pennsylvania, voted to include intelligent design in its local curriculum. And similar plans are now being considered in at least six states including Kansas, Mississippi and Arkansas. These plans include giving teachers new guidelines, and placing stickers on biology textbooks that question the scientific status of evolutionary theory.

Intelligent-design advocates have mixed reactions to the Christian right's support of their work. On the one hand, the movement is largely dependent on funding from wealthy conservative philanthropists. That, according to Meyer, is why a 1999 funding document from the Discovery Institute argued that intelligent design had "reopened

the case for a broadly theistic understanding of nature", and would eventually lay the groundwork for a series of debates and legal challenges over what should be taught in America's classrooms.

Although Meyer is willing to promote such perceptions, he concedes that they can cause problems. For intelligent-design researchers who would like to see the concept peer-reviewed and accepted by the scientific community, the politics are frustrating, and potentially dangerous. The political goals associated with intelligent design lead many scientists to reject it outright as little more than creationism in a cheap tuxedo. "Some of the policy proposals that have been made, for example the Dover case, are frankly, from our point of view, distracting," says Meyer. "We want to focus on intelligent design as an emerging research programme."

Even considered on its research merits, scientists mostly agree that intelligent design rests on shaky foundations. For one thing, Alberts points out, the concept often makes its claims based on gaps in the current body of scientific knowledge. "The whole history of science is that these gaps are always filled," he says. For example, one common argument used by intelligent-design advocates is that the bacterial flagellum, a whirling tail that some bacteria use to move around, is too complex to be explained by evolution alone. "I'm quite sure that within a decade or two we'll understand where it came from because we're sequencing more and more bacterial genomes," Alberts says. "But to give up now is totally ridiculous."

### Crisis of faith

Perhaps surprisingly, many theologians are equally upset by intelligent design. "The basic problem that I have theologically is that God's activity in the world should be hidden," says George Murphy, a Lutheran theologian, PhD physicist, and author of *The Cosmos in the Light of the Cross*. Murphy says Lutherans believe that God's primary revelation came through Jesus Christ, and many find it distasteful that additional divine fingerprints should appear in nature. Catholics, for their part, have accepted evolution based on the idea that God could still infuse the natural human form with a soul at some point in the distant past. And even the evangelical Christians who make up the backbone of intelligent design's political supporters sometimes object to its inability to prove whether Christianity is the true religion.

And yet the students listening to Cordova's lecture seem intrigued. Everyone in the room is Christian, and half are working towards degrees in science, medicine or engineering. It seems perfectly natural to them to mix science and faith. Many are also frustrated by the

## IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

exclusively secular tone of their science classes, and to these students intelligent design offers an appealing alternative that puts God squarely back in the centre of things.

Others, including Cordova himself, arrived at intelligent design from almost the opposite direction. Over a coffee earlier that day, he explains how intelligent design helped him resolve his own spiritual crisis five years ago. Since high school, Cordova had been a devout Christian, but as he studied science and engineering at George Mason, he found his faith was being eroded. "The critical thinking and precision of science began to really affect my

## IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

A decision last year at a school in Dover, Pennsylvania (left), to include intelligent design on its curriculum led residents to form a protest group (above) to combat the change.

ability to just believe something without any tangible evidence," he says. The breaking point came in 2000 when a woman from his Bible study group put her faith before her personal safety — travelling to Afghanistan as part of a covert Christian mission in a country that was, at the time, a militant Islamic theocracy. He felt unhappy accepting the promotion of such activities unless he could be sure Christianity was a true faith.

### Scientific vacuum

So Cordova turned to his scientific training in the hope of finding answers. "If I could prove even one small part of my faith through purely scientific methods that would be highly satisfying intellectually," he

says. He has since read a stack of books on cosmology and intelligent design, and has become a major advocate for the movement — representing the idea at public debates, challenging evolutionary theory in online chats and starting clubs at George Mason and several other Virginia colleges.

Cordova's story is more common than many scientists might think, according to Keith Miller, a geologist at Kansas State University in Manhattan who is an evangelical Christian. "I think a lot of students go through a period of being very conflicted about their faith, especially if they have an innate interest in science," Miller says. He knows a number of students who have fallen away from their beliefs as a result of their university experience. "They've so identified their faith with a particular view of what creation means, that it becomes an all-or-nothing kind of thing," he says. "I do think intelligent design offers an alternative, although I would argue it's not a good one."

But university lecturers are rarely able to offer students other alternatives that allow them to reconcile faith and science. Part of the problem has to do with time constraints, says Larry Rockwood, a population ecologist at George Mason. "The pressure is to work with the graduate students, do your research and teach your classes," he says. "What's the reward for working with undergraduate student clubs? Not much."

More fundamentally, most lecturers are unsure of how to handle the concerns of

## Cast out from class

Caroline Crocker says that she hadn't meant to start a controversy when she mentioned intelligent design while teaching her second-year cell-biology course at George Mason University in Fairfax, Virginia, last semester. But many of her colleagues say that the soft-spoken molecular biologist, who received a PhD in immunopharmacology from the University of Southampton, UK, has gone too far. Sitting in an empty teaching lab, Crocker tells how she has been barred by her department from teaching both evolution and intelligent design. "It's an infringement of academic freedom," she says. She is appealing the case to a grievance committee.

Crocker is one of a handful of professors nationwide who are introducing intelligent design into college-level teaching. Some, like Crocker, try to work the idea into their biology classes, but increasingly, intelligent-design advocates are teaching their material outside the science curriculum in special seminars and one-time courses, says Barbara Forrest, a philosopher at Southeastern Louisiana University in Hammond.

Those efforts meet with a mixed response from faculty members and administrators on campus. Michael Behe, an intelligent-design advocate and biochemist at Lehigh University in Bethlehem, Pennsylvania, teaches an elective first-year seminar on 'popular arguments on

evolution'. "The majority of my colleagues disagree with me," he says. "But my chairman supports my right to have my own views and argue them in a public setting."

In contrast, William Dembski, a mathematician at Baylor University in Texas and another prominent intelligent-design researcher, says that he is no longer allowed to teach on campus.

## IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

"Essentially I've had about a five-year sabbatical," he complains. Stories such as Dembski's make some intelligent-design supporters fearful of expressing their views in public. One researcher, approached by *Nature* for this article, declined to be interviewed because he did not yet have tenure.

Darwinists are divided over whether intelligent design deserves a classroom airing. Forrest says that she believes professors shouldn't be allowed to teach unsubstantiated scientific concepts to their students. "This is not a question of academic

freedom, this is a question of professional competence," she says. But Eugenie Scott, director of the National Center for Science Education in Oakland, California, which vehemently opposes teaching intelligent design in high schools, takes a different view. She thinks such discussions are more acceptable in a college environment, but believes it must be made clear to students that intelligent design is theology, not science.

Crocker hopes that she will be allowed to continue talking to students about intelligent design. Her lectures drew criticism from some and praise from others — notably, she says, her Muslim students seemed to like it. She maintains that the talks help students to think independently about ideas such as evolution. "My goal is to teach students to think for themselves," she says.

Whether and in what form her intelligent-design teachings will continue is now up to faculty members and administrators. "The university doesn't have a policy or a rule on whether certain topics should be discussed," says Daniele Struppa, a mathematician and dean of the College of Arts and Sciences at George Mason University. But, he adds, he questions whether a concept with theological underpinnings really belongs in a science course. "I'm a Buddhist," he says. "But I don't think we should teach reincarnation in biology classes."



## IMAGE UNAVAILABLE FOR COPYRIGHT REASONS



**Darwinist Eugene Scott (above) rejects intelligent design on scientific grounds whereas Lutheran George Murphy rejects it for theological reasons.**

deeply religious students, says Jo Handelsman, a plant pathologist at the University of Wisconsin at Madison. "When I talk to these students individually I don't feel it's my place to replace what their families or churches have taught them," she says. "There's a lot of confusion about where the line is, and how much it's OK to offend your students."

Scott, who is perhaps the nation's most high-profile Darwinist, is frustrated by the scientific community's inability to grapple with the issue. "The point here is that Americans don't want to be told that God had nothing to do with it," she says. "And that's the way the intelligent-design people present

evolution." Scientists need to do a better job of explaining that science makes no attempt to describe the supernatural and so has no inherent conflict with religion, she argues. "College professors need to be very aware of how they talk about things such as purpose, chance, cause and design," she says. "You should still be sensitive to the kids in your class."

Back at George Mason, Cordova is wrapping up his lecture, and planning his next steps for promoting intelligent design on campus. According to a survey he commissioned from the Campus Freethinkers — an atheist student group — 75% of students would be interested in taking a course on intelligent design if it were offered. Cordova says he hopes the poll will help convince college administrators to offer such a course. "I would love to see an intelligent-design class on one of these campuses," he says. "I don't want to indoctrinate the students; I would just like them to get to know the theory."

As for his personal future, Cordova adds that he would like to continue pursuing a career in science. Next year, he plans to apply to study cosmology at graduate school. ■

**Geoff Brumfiel is Nature's Washington physical sciences correspondent.**

## Natural divisions

From the very beginning, the purpose of Intelligent Design and Evolution Awareness (IDEA) clubs has been to facilitate debate, says Casey Luskin, who founded the first IDEA club (see picture, below) at the University of California, San Diego, in 1999.

"We want to inform everyone about all sides of the issue, so we actually invite Darwinists to the clubs to talk about natural selection," says Luskin, who now runs the IDEA Center, a small non-profit organization in San Diego that helps set up new groups on US campuses.

Evolution advocates say that researchers should be careful about how they respond to such overtures. If the request is for a public debate with an intelligent-design advocate, the best answer is 'no', argues Robert Pennock, a philosopher of science at Michigan State University in East Lansing. "A public debate is an artificial setting for getting into scientific issues," he says. "There's no way in that format to thoroughly give a scientific response, especially to a lay audience."

"A formal debate is not how we do science," agrees Eugenie Scott, director of the National Center for Science Education in Oakland, California. "But I think it's appropriate for scientists to meet with students and educate them about what the real science is saying."

That's what Victor Hutchison and his colleagues in the zoology department at the University of Oklahoma in Norman have been doing for the past few years. "We will not agree to debate the creationists publicly," he says. "But we encourage faculty members and graduate students to attend their meetings and challenge them in the discussion."

And intelligent-design supporters on campus are tolerant, more or less, of the scientists' presence. "When people remain civil, the questions that scientists ask can be illuminating," says Russell Hunter, a senior philosophy major and head of the IDEA chapter at Oklahoma. But, he adds, when scientists become too confrontational, it can have the opposite effect. "When somebody comes and gets into a yelling match, it just reinforces the beliefs of members who see the opposition as part of a political movement to make sure religion doesn't gain any ground in America," he says.



# The philosopher of photons

From meeting the Dalai Lama to national media star, Anton Zeilinger is on a mission to bring physics to a wider audience. Quirin Schiermeier listens in.

Enlightenment is hard work — especially when you mix philosophy with quantum physics. But the Dalai Lama is always keen to investigate areas of science, no matter how complex they seem. And in 1997, when he wanted to delve into the quantum world, one of the men he invited to India was Anton Zeilinger.

Based at the University of Vienna in Austria, Zeilinger is committed to explaining his work to a broad audience. As a student, he was fascinated by both opera and the mathematical beauty of quantum mechanics. And although Zeilinger, who turns 60 in May, has just revived his old student hobby of playing walk-on parts at the Vienna State Opera House, it was fundamental physics that won his heart. Over the years, his scientific achievements, as well as his communication skills, have made him a media star, at least in his native Austria.

“Explaining science to the public is hard work,” he says jovially, while clearing a spot on a visitor table that is overflowing with manuscripts. He knows that fellow researchers can be scornful of the media attention he gets. And he admits that he sometimes feels uneasy about popular portrayals of his work. When the leading German news magazine *Der Spiegel* last month featured him on its cover as ‘The Sorcerer from Vienna’, with a story full of references to time travel, parallel universes and pop culture, he was not too happy.

But we need capable translators. Quantum mechanics so often contradicts our ideas of how cause and effect determine what we see, that many physicists struggle to accept its description of the physical world. Even Einstein felt notoriously spooked by the concept of quantum entanglement, in which independent quantum particles remain mysteriously linked. So journalists appreciate Zeilinger’s ability to translate the mathematics into a working philosophy.

Zeilinger is also happy discussing the challenges of quantum mechanics — both technical and philosophical. “I strongly feel that we need to clearly tell the people what we don’t understand,” he says. For example, quantum physicists have yet to find a satisfying explanation for the randomness inherent in the quantum world. “Mysteries of that kind are among the biggest questions of the twenty-first century,” Zeilinger says.



Tough talking: Anton Zeilinger (left) discusses quantum physics and philosophy with the Dalai Lama.

When the Dalai Lama made a return visit to Zeilinger’s lab, then in Innsbruck, a year after their first meeting, he confessed to having difficulties with the philosophical implications of quantum physics, especially the role of chance and causality in nature. As the idea of determinism is central to Buddhism, the existence of purely random acts might call into question Buddhist doctrine, he said.

But Zeilinger is keen to stress that his research, and the ground-breaking quantum-optical experiments he has designed, are all explained by the known laws of physics. Less mysterious, perhaps, but equally fascinating.

## A tangled tale

When a photon of polarized laser light passes into certain types of crystal, it can generate a pair of lower-energy photons that are ‘entangled’ in terms of their respective polarization — the vertical or horizontal direction in which they oscillate. In Zeilinger’s hands, these entangled particles offer revolutionary ways of communication and computation.

One application, cryptography, has already been demonstrated, and is close to being commercially exploited. Quantum cryptography cannot prevent a spy from trying to decode a secret transmission. But the spy’s actions will irreversibly modify a secret quantum key sent from the sender to the receiver, providing cast-iron proof that an unwanted third party is listening in.

As quantum information is so easy to disturb, transmitting entangled photons over useful distances is a considerable challenge. But the technology is making good progress. Last year, for example, Zeilinger’s group

performed the first real transfer of money from a bank using quantum cryptography.

The possibilities of multiparticle entanglement are endless, Zeilinger says. Until now, entanglement experiments, such as the ‘teleportation’ of a complete quantum state from one particle to another, have used up to four particles (usually photons). If it were possible to ‘connect’ even more particles and maintain high-quality entanglement over greater distances, this may open up entire new worlds of quantum communication, he says.

Zeilinger dreams of using satellites for quantum communication between any two spots on the globe. Although transmitting entangled photons on the ground is limited to distances of less than 20 kilometres, in space it is much easier. A feasibility study for supplying satellites with high-precision lasers to produce entangled photons and send them to Earth has already been made. Zeilinger hopes that the European Space Agency will fund a payload of such components to the International Space Station by 2010.

The ultimate goal might be to connect a network of future quantum computers. Zeilinger’s group has recently demonstrated the first one-way quantum computer, in which an entangled multi-photon state starts out with all of the possible calculation results, and proceeds by irreversible steps to reach the final answer (P. Walther *et al.* *Nature* **434**, 169–176; 2005). But a future ‘Internet’ based on the laws of quantum physics is currently pure science fiction.

Something else for the Dalai Lama, and the rest of us, to meditate over.

Quirin Schiermeier is *Nature*’s German correspondent.

# A drug is effective if better than a harmless control

Valid trials can still be held, as with HIVNET 012, when ethics rules out a placebo group.

**Sir**—In his Correspondence letter “HIV drug remains unproven without placebo trial” (*Nature* **434**, 137; 2005), Valendar Turner asks on what basis it can be claimed that the single-dose nevirapine regimen is effective in reducing mother-to-infant HIV transmission.

He points out that the placebo arm was dropped in the HIVNET 012 trial (of which we were team members) and notes the high variability in reported historical HIV transmission rates.

But he ignores the valid comparison made by the HIVNET 012 trial — whose conclusions were recently accepted by a US Institute of Medicine panel — between nevirapine and another drug, zidovudine, which had itself been tested against a placebo in other studies starting at least four weeks before labour.

The HIVNET 012 trial in Uganda originally placed HIV-infected pregnant women and newborns in three groups, randomly selected, to receive either the nevirapine regimen (a single dose of nevirapine to the mother during labour and a single dose to the baby shortly after birth), or a very short course of zidovudine from the onset of labour through the first week of the baby's life, or a placebo.

As noted by Turner, the placebo arm of HIVNET 012 was dropped shortly after the study began, when a Thai study showed that a short course of zidovudine, given

for the last four weeks of pregnancy and during labour, reduced transmission by 50% compared with placebo (N. Shaffer *et al.* *Lancet* **353**, 773–780; 1999).

Women on HIVNET 012 continued to be enrolled in the two active drug arms, following recommendations — made jointly by the US Centers for Disease Control and Prevention, the National Institutes of Health, the French National Agency for AIDS Research, and the Joint United Nations Programme on HIV/AIDS — to drop the placebo arm from all ongoing perinatal HIV trials.

Once the placebo arm was dropped, the redesigned trial compared the mother-to-infant transmission rates for nevirapine and on zidovudine within the same trial, rather than making a comparison with historical transmission rates, as Turner implies.

When one compares two drugs, if the test and control drugs are found to be similar in efficacy, it is not possible to know whether the two drugs are equally effective or equally ineffective — unless there is clear evidence that the control drug is effective. This is why use of placebos is advocated in settings where the active control has not been established to have efficacy that is clinically and statistically significant.

But when an experimental drug is found to be superior to a control that itself

is not harmful (thus replacing a placebo), the effectiveness of the experimental drug is thereby established.

In the HIVNET 012 trial, a very short course of zidovudine was used as the active control. Many studies had already reported that longer zidovudine regimens were more effective than placebos.

Of the HIVNET 012 patients who gave birth before discontinuation of the placebo arm, the transmission rate at 6–8 weeks of age was 37% in the placebo group, compared with 28% in the zidovudine group and 7% in the nevirapine group. (The 6–8-week transmission rates at the end of the trial were 20% for zidovudine and 11.8% for nevirapine.)

Most people would conclude that zidovudine does not increase the risk of transmission. Hence, our finding that single-dose nevirapine was significantly more effective in preventing HIV transmission than a very short course of zidovudine justifies the conclusion that the HIVNET 012 nevirapine regimen is more effective than nothing in preventing mother-to-infant HIV transmission.

**Brooks Jackson\*, Thomas Fleming†**

\**Johns Hopkins School of Medicine, 600 North Wolfe Street, Baltimore, Maryland 21287, USA*

†*Department of Biostatistics, University of Washington, Box 357232, Seattle, Washington 98195, USA*

## DNA barcoding does not compete with taxonomy

**Sir**—M. C. Ebach and C. Holdrege, in Correspondence (“DNA barcoding is no substitute for taxonomy” *Nature* **434**, 697; 2005), express some key misunderstandings regarding what a comprehensive DNA barcoding programme would — and especially would not — do. DNA barcoding projects are already achieving positive results, albeit on a relatively small scale. If implemented globally, DNA barcoding would benefit, not compromise, taxonomic science.

It is important to note that it does not seek to replace the linnaean system of classification, and thus differs fundamentally from proposals to create a new taxonomic system based solely on DNA. With only 15% of the estimated 10 million species described, large-scale DNA barcoding could be used to highlight probable new species within previously unstudied taxa. However, these species-in-waiting would not be named solely with a DNA barcode:

they would be given linnaean names based on the study of curated voucher specimens, high-resolution digital images, collection locality data and other information.

Perhaps the most unfortunate misunderstanding is that DNA barcoding competes with taxonomy for funding. Existing DNA barcoding networks have been funded by agencies that do not have a tradition of supporting taxonomic work. A global DNA barcoding initiative would be a ‘big science’ programme, and as such would compete for priority with projects of similar scale from physics, medicine and genomics — not taxonomy. Moreover, the bulk of such funds would ultimately be directed toward the collection and curation of specimens, not DNA sequencing.

Rather than draining support from taxonomy, the DNA barcoding initiative has the potential to inject significant new funding into museums, herbaria and individual taxonomy labs.

**T. Ryan Gregory**

*Department of Integrative Biology, University of Guelph, Guelph, Ontario N1G 2W1, Canada*

## Nice planet, shame about the human race

**Sir**—Inspired by Stephen Baxter's Futures story “Under martian ice” (*Nature* **433**, 668; 2005), we began to discuss the Fermi paradox: that if aliens exist, they would have visited everywhere by now, including Earth. Careful consideration led us to conclude that if they were intelligent, they would not visit this planet.

Thus — and in opposition to the anthropic principle, which argues that the Universe is the way it is because we are here to observe it — we propose the misanthropic principle as the resolution of the puzzle.

**Randall D. Kamien, Madhuri Kaul**

*Department of Physics and Astronomy, 209 South 33rd Street, University of Pennsylvania, Philadelphia 19104, USA*

### correspondence

Contributions to Correspondence may be submitted to [corres@nature.com](mailto:corres@nature.com).



# Seeing sense

A look back at work that established the link between eye and brain.

## Brain and Visual Perception: The Story of a 25-Year Collaboration

by David H. Hubel & Torsten N. Wiesel  
Oxford University Press: 2004. 738 pp.  
\$49.50, £29.99

Charles G. Gross

In 1959, two postdoctoral fellows, a Canadian and a Swede working together at Johns Hopkins University, made an 'accidental discovery'. They found that the image of an edge of a glass slide activated cells in the region of a cat's brain known as the striate cortex. Over the next 25 years, David Hubel and Torsten Wiesel expanded their observation into some of the most important developments in understanding sensory physiology, sensory psychology and the functional architecture of the cerebral cortex since E. A. Adrian initiated modern sensory neurophysiology some 75 years ago.

Hubel and Wiesel outlined a scheme for the hierarchical processing of visual information that became the frame for all subsequent work in visual neurophysiology, as well as a model for other sensory systems. They provided the first possibility of a bridge from the eye to the cognitive science of pattern recognition. Their work included the first experimental demonstration that normal experience sculpts the anatomy and physiology of the brain. They also carried out the first experimental examination of the relative role of experience and innate wiring in the development of the neural mechanisms of perception and behaviour.

Almost uniquely among basic neuroscience researchers, Hubel and Wiesel made a discovery that was immediately transferred to the clinic, in this case saving binocular vision for children born with strabismus ('crossed' or 'wall' eyes). Their experiments were so simple and elegant that they had no use for computers or elaborate apparatus, and their results were so clean that no statistics were needed. They did all this and more, essentially with their own hands and without the help of graduate students, postdocs or an army of technicians. They received a Nobel prize for their work in 1981.

The bulk of this book consists of reproductions of 28 of Hubel and Wiesel's most

important papers. The book begins with autobiographical essays by both men. The rest of the book is written by Hubel, but the 'we' that he uses is very much a collaborative rather than a royal one.

The autobiographies are followed by several short background chapters. One, too brief and forgiving, describes the state of cortical physiology when Hubel and Wiesel began their work. The retinotopic organization of striate cortex was understood from mapping studies in cats and monkeys done by Wade Marshall and S. A. Talbot in the 1940s, but nothing was known of the response properties of individual striate-cortex neurons.

In 1953 there were two demonstrations concerning cells in the frog's retina that could analyse complex form, one by Horace Barlow and one by Steve Kuffler, who was to become Hubel and Wiesel's postdoctoral adviser. These studies provided the starting point for Hubel and Wiesel's investigations of the cat's visual cortex in 1959, and probably sensitized the pair to the importance of the observation described in the opening of this review. Serendipity requires a prepared mind.

At about the same time, the work of Barlow and Kuffler was extended by Jerry Lettvin and Humberto Maturana in their paper 'What the frog's eye tells the frog's



Recognition: David Hubel (left) and Torsten Wiesel won a Nobel prize for their work in 1981.

Other recording studies of striate cortex at this time usually used electrical stimulation of the optic nerve, rather than discrete visual stimuli.

By the 1950s, techniques for single-neuron recording in the striate cortex had been established by groups in Montreal in Canada and Freiburg in Germany. Yet both groups usually used only diffuse illumination as the visual stimulus, and this elicited only feeble responses. Eventually the Freiburg group started using an elaborate apparatus that could move a grid across the visual field, but only in a vertical orientation, making the discovery of orientation selectivity impossible. The only responses they found were variations of the 'on', 'off' and 'on/off' responses described by the Nobel laureate H. Keffer Hartline some 40 years earlier.

Hubel and Wiesel were impressed when Lettvin showed them his complex frog neurons. The pair realized they "obviously had much in common with Lettvin and Maturana, especially our exploratory approach and freedom from complex apparatus, hypothesis, and so on. Jerry and Humberto carried their unusual approach to extremes, perhaps because they lacked an adviser like Steve who insisted we make at least a few measurements for the sake of scientific respectability."

The remaining background chapters in this book are largely a glowing tribute to Kuffler, Hubel and Wiesel's mentor, friend and protector, who died in 1980. He brought them together at Johns Hopkins, encouraged them in every way, and took them to Harvard. Beyond this, Kuffler provided a lifelong model of doing science with your own hands,

writing and rewriting over and over, and leaving your graduate students alone.

Perhaps the most interesting parts of the book are those that surround each paper or set of related papers. The forewords describe why they did each experiment. Then they describe what they did right and wrong, and comment a little (actually, very little) on subsequent developments in the field. Best of all, these parts, and the introductory chapters, are peppered with incidental remarks on how (and how not) to do science.

For example, on grants they say that much of their research "can be described as a massive fishing expedition, an expression commonly used by study sections to disparage bad grant requests". Their research, they explain, was seldom hypothesis driven. "But the lack of a hypothesis need not necessarily prevent one from catching big fish."

On computation: an example of the "illnesses" that can afflict science is an increase in "a theory sometimes called computation ... molecular biology, which we regard as more successful as a science than our field, seems largely to have avoided being beset with computation. In *The Molecular Biology of the Gene* I look in vain for equations."

On doing experiments, they write: "Unlike much of today's science, in which the actual work is done by technicians or graduate students ... it is we who get to do the experiments."

Hubel says he saves time by reading as little as possible in his field: "Reading most papers today is like eating sawdust." He also says the pair benefited from their "refusal to waste time bothering with measuring intensities, rates of movement, and so on, or to spend time drawing graphs or histograms."

On statistics: "We could hardly get excited about an effect so feeble as to require statistics for its demonstration." And on failing to notice the directional properties of MT cells despite recording from about 200 of them: "We were lazy and not very bright."

There are two subjects on which I would have liked to see more. The first is their students: who were they, what did they work on, and how were they mentored? Perhaps they were just left alone. There is more about how Kuffler helped the duo than how they treated their own students.

The other subject is how the two collaborators actually worked together. One disagreement is obliquely referred to without explaining what it was. They speak of bulging files of 30 years of experimental protocols, a folder for each experiment. It would have been instructive to include some examples of these protocols, even if it meant fewer reprinted papers in the volume. But perhaps those files are for the historians, rather than an autobiographical volume.

Charles G. Gross is in the Department of Psychology, Princeton University, Green Hall, Princeton, New Jersey 08544-1010, USA.



Crunch time: large numbers of people used to do the calculations now done by electronic computers.

## Recovered history

### When Computers Were Human

by David Alan Grier

Princeton University Press: 2005. 412 pp.

\$35, £22.95

Jon Agar

*We work from morn till night,  
For computing is our duty;  
We're faithful and polite,  
And our record book's a beauty;  
With Crelle and Gauss, Chauvenet  
and Peirce,*

*We labor hard all day;  
We add, subtract, multiply and divide,  
And we never have time to play.*

(from *The Observatory Pinafore* by Winslow Upton, 1879)

Human computers certainly did work hard all day, and had the aches and pains to show for it: elbow joints inflamed from cranking calculator handles, or fingers and thumbs cramped from pencilling figure after figure on to graph paper. For two centuries the modern scientific enterprise was built on their efforts. Yet every time a logarithm was looked up, or the value of a Bessel function checked, a debt was incurred that was rarely acknowledged. And now the human computer has vanished from history.

People do not disappear from our collective memory by accident. Forgetting is not a passive process: people are forgotten for a reason. Two decades ago, the sociologist Steven Shapin noted pointedly that the lab technician, although essential to making an experiment work, rarely appeared in accounts of successful scientific work. But if something went wrong, the finger was pointed at human interference. Erasing the human hand was part of the means by which

an experiment was seen to reveal aspects of nature, rather than aspects of society.

David Alan Grier's recovery of the wonderfully rich story of human computers not only allows us to repay a debt, but also to ask why human computers were made to disappear in the first place. They were drawn from the margins of the scientific establishment. Many were women. Nicole-Reine Lepaute, for example, the well-to-do wife of a royal clockmaker, was one of a trio — said to be the first astronomers to divide the labour of scientific calculation — who calculated the orbit of Halley's comet in time for its perihelion of 1758. Many later female computers came from much poorer backgrounds.

Male computers, too, were relatively disadvantaged. Many of the computers employed in Gaspard de Prony's Bureau du Cadastre, a factory of calculation, were hairdressers and wigmakers who had fallen on hard times following the French Revolution. Often the reason was financial: women, boys and immigrants made for cheap and willing labour. It was even said of Harvard Observatory, run by the relatively progressive Edward C. Pickering and staffed by college graduates, that the "computers are largely women who can be got to work for next to nothing".

By the Second World War, the heyday of human computers, they were routinely referred to as 'girls'. Indeed, managers calculated 'girl-years' of effort, and even defined the 'kilogirl', a unit for measuring work. But, as Grier emphasizes, "even at this date, computing was not the sole domain of women. It was really the job of the dispossessed, the opportunity granted to those who lacked the financial or societal standing to pursue a scientific career." For computing offered people from the margins an entry into the scientific world — an opportunity that they often grasped with both hands.

It is notoriously difficult to recover details

of the lives of ordinary people: they rarely leave their words for posterity, and can be glimpsed only through the writings of their social superiors. Many more words have been written about Napoleon than about his foot soldiers. Indeed, it would take the imaginative power of a Tolstoy to reach back and reanimate the human computers labouring in the calculating offices of the eighteenth and nineteenth centuries. Despite Grier's industry, many still remain anonymous.

But Grier triumphantly achieves his aim when discussing the twentieth-century human computer, as many are alive to tell their tales. Take, for example, the life of Gertrude Blanch. Born Gittel Kaimowitz in Poland, and educated well, she fled the pogroms with her family. In the United States,

Gertrude struggled for years to find employment that matched her mathematical ability. But a chance meeting in 1937 with the director of the Mathematical Tables Project, Arnold Lowan, who shared her background, presented her with an opportunity that she gladly grasped. By 1940, the project was the largest scientific computing organization in the United States, and Blanch was one of its organizers. Grier traces in detail how Blanch's later career was blighted by the FBI's suspicions that she was a communist sympathizer.

Blanch's story is told with Grier's characteristic verve. But why did she, and her fellow human computers, disappear from history? Partly it was a familiar case of science erasing the traces of its human creation. But also

the human computers suffered from the need to provide their successor with a suitable genealogy. At the seminal Moore School summer classes, where the team that built the ENIAC (the Electronic Numerical Integrator and Computer) spoke of new stored-program computers, listeners "heard a somewhat fanciful history of calculating devices that ignored the contributions of [human] computers". This "was an attempt to build a distinguished lineage for the electronic computing machine, a pedigree that ignored the influence of commerce and the hard labour of human computers."

Left undisturbed, victors write history. ■

Jon Agar is in the Department of History and Philosophy of Science, University of Cambridge, Cambridge CB2 3RH, UK.

## Theatre

### Artistic differences

#### Phallacy

written by Carl Djerassi

Performed at the New End Theatre, London, until 14 May 2005

#### Robin Clark

*Phallacy* is an intriguing play, well acted, fast moving and embracing a host of questions and human situations that are rarely touched upon in modern theatre. It was written by the distinguished synthetic chemist Carl Djerassi, who is best known for his role in the creation of the contraceptive pill. In recent years, Djerassi has turned his attention to writing scientifically based plays, including *Calculus* and (with Roald Hoffmann) *Oxygen*. *Phallacy* derives not only from his interests in science, but also from those as an art collector, and raises questions that arise at the interface of the arts and the sciences.

The play is based on the actual case of a revered life-sized bronze of a young man held in the Kunsthistorisches Museum, Vienna. It had been considered for several centuries to be a Roman original, its image even appearing on an Austrian postage stamp. However, scientific analysis in 1986 — specifically, thermoluminescence studies — revealed it to be a Renaissance cast. In Djerassi's play, this result opens the door to questions about how an object is devalued in such circumstances (really a question to do with market forces) and, more importantly, to a clash between professional reputations in the arts and the sciences.

The key players in this clash are a leading art historian, Regina Leitner-Opfermann

(convincingly played by Karen Archer), and a professor of chemistry, Rex Stolz (played equally well by Jack Klaff), who had been invited to date the statue scientifically. She has written a 345-page book on the statue, describing every art-historical aspect of it in great detail. But her book lacks reference to any scientific analysis that might have a bearing on whether or not the statue dates to the correct period; the word 'thermoluminescence' doesn't even appear in the index.

The art historian — supposedly in search of 'artistic truth' — is a 'true believer', and can see her work and reputation being undermined by the chemist, who is concerned with

been insufficient communication. There is no 'artistic truth' or 'scientific truth' as such, but a presumed universal truth that requires for its establishment much more communication between the arts and sciences than is currently the case. It is certainly the scientists who are making the running here: many do read arts literature, whereas it is rare for art historians to read scientific literature, even that bearing upon art and artefacts.

Djerassi neatly highlights the importance of, and urgent need for, the scientific investigation of works of art. Such studies have recently called into question a host of heavily entrenched opinions that lack scientific credibility — notably concerning the Turin Shroud, the Vinland Map and certain Vermeer paintings.

A speculative sub-plot as to the likely provenance of the statue adds colour and entertainment value to the play without really impinging upon the main plot and debate. Such sub-plots are less appropriate in documentary-type productions of real cases of art evaluation, such as Nova Productions' *The Viking Deception*, which was recently shown on US television and in which I was interviewed about the analysis of the ink used on the Vinland Map. Much unnecessary space was devoted to the question of who might have drawn the Vinland

Map, given that it wasn't the Vikings.

The supporting cast of *Phallacy* were excellent, notably the art historian's assistant Emma Finger (played by Lucy Liemann) and the professor's assistant Otto Ellenbogen (Hamish Clark). They all contributed to an appealing and thought-provoking new production. ■

Robin Clark is at Christopher Ingold Laboratories, University College London, 20 Gordon Street, London WC1H 0AJ, UK.



Supporting cast: Hamish Clark (left) and Lucy Liemann play figures on either side of a debate about the origins of a bronze in the play *Phallacy*.

the facts, the scientific truth of the situation. He is in no doubt as to the correctness of his conclusions, which point to the statue being a sixteenth-century cast of hollowed bronze, which has a different trace metal content from that of Roman bronze and contains a far higher percentage of nickel.

The play is presented as a clash between incompatible cultures. I see instead a clash of the reputations of people working in different disciplines between which there has



# Predicting with unpredictability

Random numbers: from stone casting to computers to radioactive decay, the generation of random sequences has always preoccupied mankind.

Gianpietro Malescio

**M**aking predictions is one of the main goals of science. Traditionally this implies writing down, and solving, the equations governing the system under investigation. When this method proves impossible we often turn to a stochastic approach. The term 'stochastic' encompasses a variety of techniques that are based on a common feature: using unpredictable entities — random numbers — to make predictions possible.

The origins of stochastic simulation can be traced to an experiment performed in the eighteenth century by Georges Louis Leclerc, Comte de Buffon. Leclerc repeatedly tossed a needle at random on to a board ruled with parallel straight lines. From his observations, he derived the probability that the needle would intersect a line. Subsequently, Pierre Simon de Laplace saw in this experiment a way to obtain a statistical estimate of pi.

Later, the advent of mechanical calculating machines allowed numerical 'experiments' such as that performed in 1901 by William Thomson (Lord Kelvin) to demonstrate the equipartition theorem of the internal energy of a gas. Enrico Fermi was probably the first to apply statistical sampling to research problems, while studying neutron diffusion in the early 1930s. During their work on the Manhattan Project, Stanislaw Ulam, John von Neumann and Nicholas Metropolis rediscovered Fermi's method. They established the use of random numbers as a formal methodology, generating the 'Monte Carlo method' — named after the city famous for its gambling facilities. Today, stochastic simulation is used to study a wide variety of problems (many of which are not at all probabilistic), ranging from the economy to medicine, and from traffic flow to biochemistry or the physics of matter.

When the temporal evolution of a system cannot be studied by traditional means, random numbers can be used to generate an 'alternative' evolution. Starting with a possible configuration, small, random changes are introduced to generate a new arrangement: whenever this is more stable than the previous one, it replaces it, usually until the most stable configuration is reached. Randomness cannot tell us where the system likes to go, but allows the next best thing: exploration of the space of the configurations while avoiding any bias that might exclude the region of the

IMAGE  
UNAVAILABLE  
FOR COPYRIGHT  
REASONS

Predicting the unpredictable: casting rune stones has been used as a divination tool for more than 2,000 years.

possible solution. If we are able to guess the probability distribution of the configurations, then instead of conducting a uniform random search we can perform an 'importance' sampling, focusing our search on where the solution is more likely to be found.

Optimization problems are often solved using stochastic algorithms that mimic biological evolution. Although it may sound vaguely unpleasant, we come from a random search. In nature, new genetic variants are introduced through random changes (mutations) in the genetic pool while additional variability is provided by the random mixing of parent genes (by recombination). Randomness allows organisms to explore new 'designs' which the environment checks for fitness, selecting those most suited to survival. But the optimal solution is not found once and for ever. A continually changing environment means evolution is an ongoing process; it does not produce the 'perfect' organism, but rather a dynamic balance of myriad organisms within an ecosystem.

Generating true randomness is a challenging task. Early attempts at stochastic simulation produced samples through processes such as dice tosses or card draws. These phenomena in principle obey newtonian mechanics, but in practice evolve unpredictably owing to their chaotic dynamics. Computers have made it much easier to produce large numbers of samples. They cannot, however, generate true random numbers.

As Gregory Chaitin and Andrei Nikolae-vich Kolmogorov first pointed out in the 1960s, a random sequence is algorithmically incompressible; that is, it cannot be represented with a description shorter than the sequence itself. A string of 'random' numbers generated by a computer can be easily compressed since it can be described by the

algorithm that generates the sequence. Although computers can produce pseudorandom numbers that are virtually indistinguishable from true random numbers, all such sequences eventually show correlations, and fall into a repeating pattern.

To circumvent these problems, scientists are studying physical devices that can produce efficiently random numbers. A variety of phenomena have been considered, including the detection of atmospheric chaos by a noisy radio receiver, the timing of radioactive decays, and the beaming of photons at a semi-transparent mirror. In principle, random numbers captured in the real

world are superior to those generated by computers. In practice, sequences of pseudorandom numbers may seem more convincingly random than those produced by hardware generators.

Devising random-number generating physical processes that are precisely balanced is not easy. For example, coin tossing — a seeming paradigm of randomness — is actually subtly biased owing to the mass imbalance derived from the different designs on the two sides of the coin. Even if carefully engineered to minimize bias and other artefacts, hardware generators need 'whitening' algorithms to treat residual correlations, and statistical tests to detect possible failures in randomness. This can be checked only by using computers, which are notoriously unable to produce true random numbers! Thus, more reliable and efficient generators of random numbers cannot rely solely on either hardware or software, but should employ a combination of both.

Creating perfect disorder, eliminating even the slightest bias or correlation, is as difficult a task as maintaining perfect order. Nevertheless, the search for pure noise will go on because the usefulness of random numbers for making predictions resides in their complete unpredictability. As Chaitin puts it: "The world consists of the tension between order and chaos. When simulating physical phenomena, order is supplied by the laws of physics and chaos is supplied by random numbers."

Gianpietro Malescio is in the Department of Physics, University of Messina, 98166 Messina, Italy.

## FURTHER READING

Metropolis, N. *Los Alamos Sci.* **15**, 125–130 (1987).

Chaitin, G. J. *Sci. Am.* **232**, 47–52 (1975).

Calude, C. S. & Chaitin, G. J. *Nature* **400**, 319–320 (1999).

# A certain flare

Davide Lazzati

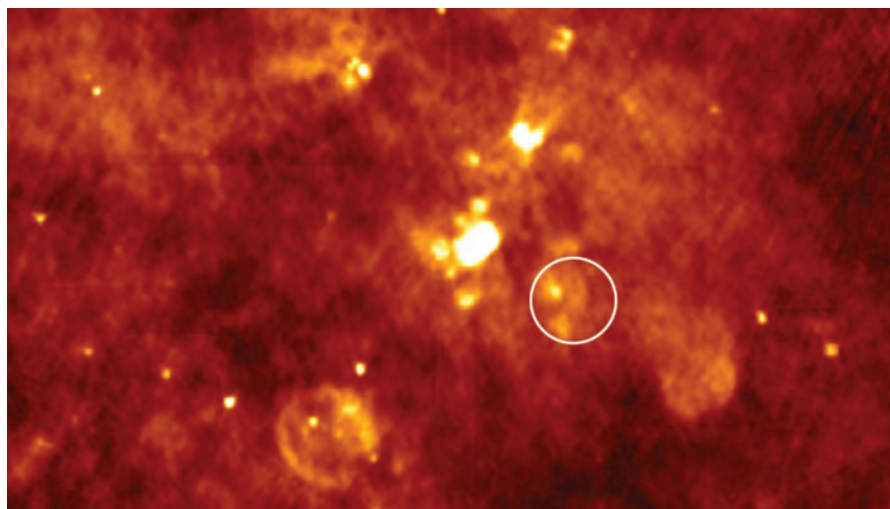
**Giant flashes from soft  $\gamma$ -ray repeaters are spectacular but rare events — only three have ever been observed in our Galaxy. The suspicion is that we have been missing some from farther afield.**

On 27 December 2004, virtually all of the  $\gamma$ -ray detectors in orbit were triggered by the brightest flash of  $\gamma$ -rays ever seen. Two similar flares had previously been detected from different sources of the same class during 30 years of observations<sup>1,2</sup> — on 5 March 1979 and 27 August 1998. The 2004 flare, however, must be regarded as unique: it outshone both the preceding events by two orders of magnitude, releasing in its first fraction of a second as much energy as the Sun releases in a quarter of a million years. Five papers<sup>3–7</sup> in this issue provide an observational overview of this exceptional event.

The source of the outburst is known as SGR 1806–20, a ‘soft  $\gamma$ -ray repeater’ (SGR) lying in our Galaxy at an estimated distance of 15 kiloparsecs (almost 50,000 light years) from the Solar System (Fig. 1). An SGR is an extremely highly magnetized neutron star<sup>8</sup>, or ‘magnetar’, that produces recurrent bursts of low-energy (‘soft’)  $\gamma$ -rays — that is, high-energy photons, or electromagnetic radiation. The flare from SGR 1806–20 was characterized by an initial spike that lasted less than a second and contained most of the energy of the burst<sup>3,4</sup> as well as the highest-energy, or ‘hardest’, photons. This spike of the flare was followed by an exponential tail with a duration of some 400 seconds, oscillating with the period (7.56 seconds) at which SGR 1806–20 is known — from measurements of its much dimmer X-ray emission during quiescence — to rotate.

The characteristics of the SGR 1806–20 flare can be explained as the outcome of a readjustment of the huge magnetic field — up to  $10^{15}$  times stronger than that at Earth’s surface — anchored to what is a relatively young (about 5,000-year-old) neutron star<sup>8</sup>. Such a readjustment releases a sizeable fraction of the internal energy of the field, stored in a hot ‘plasma’ of radiation and electron–positron pairs, and generates the bright initial spike of the flare.

The flux of photons in the spike of SGR 1806–20 was so large that it saturated most detectors<sup>3</sup>, making it difficult to characterize its properties. Terasawa *et al.*<sup>5</sup> (page 1110), however, report an oscillatory modulation, with a period of around 60 milliseconds, in the number of photons detected in the spike. They suggest that the periodicity of these ‘humps’ in the flare’s profile indicates repeated injections of energy into



**Figure 1 Site unseen.** A wide-field view of the area around SGR 1806–20 (at the centre of the white circle) before the colossal  $\gamma$ -frequency flare of 27 December 2004, from radio frequency measurements. At this point SGR 1806–20 was still ‘radio quiet’; an intense radio nebula emanating from the neutron star was only observed days after the first  $\gamma$ -ray burst<sup>6,7</sup>.

the system, the origin of which is unclear.

The pulsed emission in the flare’s tail comes from a fraction of the plasma that initially remains confined by the magnetic field lines anchored to the neutron star. As the star rotates, it produces a ‘lighthouse’ effect, resulting in the periodic oscillations in brightness.

On a timescale of days after the initial flare, X-ray<sup>9</sup> and radio emissions<sup>6,7</sup> originating from SGR 1806–20 were also observed. Gaensler *et al.*<sup>6</sup> (page 1104) report that the expanding radio nebula had a luminosity more than 500 times greater than the only other similar object detected (after the flare of August 1998). In some ways the afterglow was similar to that of long  $\gamma$ -ray bursts (GRBs). GRBs are another class of bright  $\gamma$ -ray flashes that are known to originate from galaxies on the edge of the visible Universe. Long GRBs last for tens of seconds and are followed by a dimmer emission tail lasting weeks to months. Cameron *et al.*<sup>7</sup> (page 1112), however, point out general problems in reconciling the spectrum and light curve of the SGR 1806–20 radio emission with standard models of long GRBs.

The most striking property of the SGR 1806–20 flare is its extraordinary luminosity. This raises the question of whether this giant flare and its two predecessors<sup>1,2</sup> are related to a class of mysterious short GRBs, detected in large numbers by BATSE (the

Burst and Transient Source Experiment), a soft  $\gamma$ -ray detector that flew on board NASA’s Compton observatory in the 1990s. No afterglow emission has so far been identified from these GRBs. However, if we rescale the properties of the SGR 1806–20 flare to a distance of several megaparsecs (around a thousand times farther away), an instrument such as BATSE would indeed have seen only the initial bright spike of the event, with a timescale of several hundred milliseconds and a spectrum consisting almost entirely of hard photons. Such a description matches that of the short GRBs detected. Hurley *et al.*<sup>3</sup> (page 1098) estimate that the rate of giant flares that should have been detectable by BATSE is about 30 per year, which could account for up to 40% of the short GRBs actually found.

There are three ways to identify the presence of SGR flares in the BATSE catalogue of short bursts. First, they should be much closer than cosmological GRBs, and therefore associated with bright galaxies. The position in the sky is known with the required accuracy for only five short GRBs; but no bright host galaxy can be associated with any of these five<sup>10</sup>, limiting the possible proportion of SGR flares to less than 20% in the BATSE short-burst catalogue. Second, SGRs should produce a periodic signal in the 200 seconds following the burst; a search for such a signal has so far been unsuccessful. Third, SGR candidates in the BATSE catalogue can be

UNIV. HAWAII

identified by their spectrum. SGR flares typically emit a thermal 'black-body' spectrum (indicative of an optically thick medium), whereas GRBs are characterized by broad power-law spectra (emitted by optically thin material) with the radiation spread over many orders of magnitude in frequency. The SGR 1806–20 flare had an average temperature of  $2 \times 10^9$  kelvin, which should be easily identifiable among non-thermal GRB spectra; but a search for high-temperature thermal spectra performed over about 100 BATSE short GRBs has been unsuccessful.

These factors constrain the percentage of possible SGR flares among short GRBs to 5% at most, close to an order of magnitude less than expected. So, where are the extragalactic SGR flares?

Several factors may explain the apparent discrepancy. First, the distance to SGR 1806–20 could be inaccurate. If the flare were at only half the distance assumed, its luminosity would be reduced by a factor of four, limiting the volume of space in which extragalactic SGRs could be detected by a factor of eight. A possible smaller distance to SGR 1806–20 has been proposed. An upper limit of 9.8 kiloparsecs, two-thirds of the previous working assumption<sup>11</sup> of 15 kiloparsecs, is suggested by Cameron *et al.*<sup>7</sup>

Another factor in the calculation is the rate of occurrence of giant flares in our Galaxy. SGR 1806–20 is the only flare of its scale to be seen in 30 years of observations, and this is our best estimate of how often they occur. Of course, we may merely have been lucky, and the true rate could be lower: we might just happen to live in an epoch when one of these flares went off randomly.

Whatever the answers to these questions, the hunt for extragalactic flares is on. In their analysis, Palmer *et al.*<sup>4</sup> (page 1107) use data from NASA's Swift satellite, which was launched in 2004 expressly to solve the GRB mystery, and the rapid follow-up capabilities of this mission should facilitate further discoveries. The importance of such detections for our understanding of an extreme phenomenon that is represented so far only by the example of 27 December 2004 cannot be overestimated.

Davide Lazzati is at JILA, University of Colorado at Boulder, 440 UCB, Boulder, Colorado 80309-0440, USA.

e-mail: lazzati@colorado.edu

1. Mazets, E. P. *et al.* *Nature* **282**, 587–589 (1979).
2. Hurley, K. *et al.* *Nature* **397**, 41–43 (1998).
3. Hurley, K. *et al.* *Nature* **434**, 1098–1103 (2005).
4. Palmer, D. M. *et al.* *Nature* **434**, 1107–1109 (2005).
5. Terasawa, T. *et al.* *Nature* **434**, 1110–1111 (2005).
6. Gaensler, B. M. *et al.* *Nature* **434**, 1104–1106 (2005).
7. Cameron, P. B. *et al.* *Nature* **434**, 1112–1115 (2005).
8. Thompson, C. & Duncan, R. C. *Mon. Not. R. Astron. Soc.* **275**, 255–300 (1995).
9. Mereghetti, S. *et al.* *Astrophys. J. Lett.* (submitted); preprint available at [www.arxiv.org/astro-ph/0502577](http://www.arxiv.org/astro-ph/0502577) (2005).
10. Nakar, E., Gal-Yam, A., Piran, T. & Fox, D. B. preprint available at [www.arxiv.org/astro-ph/0502148](http://www.arxiv.org/astro-ph/0502148) (2005).
11. Corbel, S. & Eikenberry, S. S. *Astron. Astrophys.* **419**, 191–201 (2004).

## Evolutionary biology

# Animal roots and shoots

Martin Jones and Mark Blaxter

DNA sequence data from neglected animal groups support a controversial hypothesis of deep evolutionary history. Inferring that history using only whole-genome sequences can evidently be misleading.

Despite the comforting certainty of textbooks and 150 years of argument, the true relationships of the major groups (phyla) of animals remain contentious. In the late 1990s, a series of controversial papers used molecular evidence to propose a radical rearrangement of animal phyla<sup>1–3</sup>. Subsequently, analyses of whole-genome sequences from a few species showed strong, apparently conclusive, support for an older view<sup>4–6</sup>. Philippe *et al.*, writing in *Molecular Biology and Evolution*<sup>7</sup>, now provide evidence from expanded data sets that supports the newer evolutionary tree, and also show why whole-genome data sets can lead phylogeneticists seriously astray.

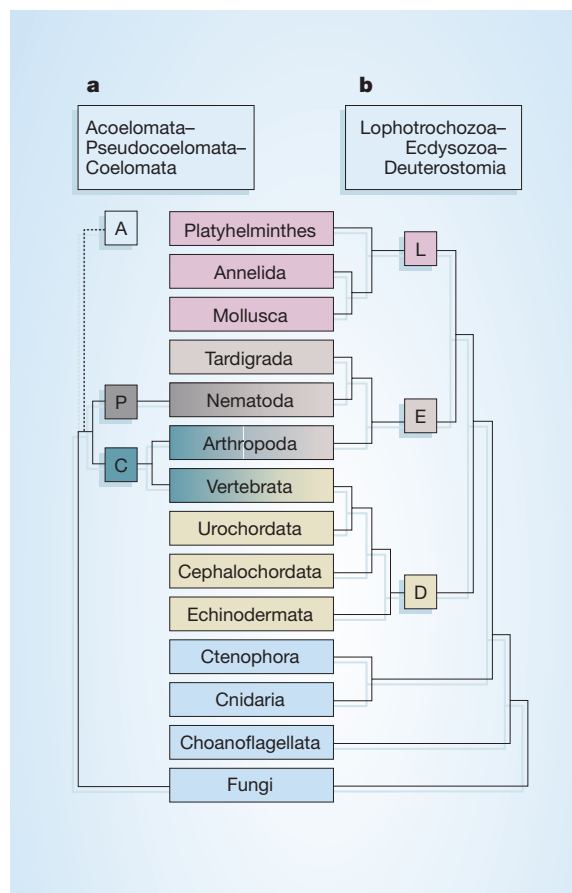
Traditional trees group together phyla of bilaterally symmetrical animals that possess a body cavity lined with mesodermal tissue, the coelom (for example, the human pleural cavity), as Coelomata. Those without a true coelom are classified as Acoelomata (no coelom) and Pseudocoelomata (a body cavity not lined by mesoderm). We call this

tree the A–P–C hypothesis. Under A–P–C, humans are more closely related to the fruit-fly *Drosophila melanogaster* than either is to the nematode roundworm *Caenorhabditis elegans*<sup>5,6</sup> (Fig. 1).

In contrast, the new trees<sup>1–3,7</sup> suggest that the basic division in animals is between the Protostomia and Deuterostomia (a distinction based on the origin of the mouth during embryo formation). Humans are deuterostomes, but because flies and nematodes are both protostomes they are more closely related to each other than either is to humans. The Protostomia can be divided into two 'superphyla': Ecdysozoa (animals that undergo ecdysis or moulting, including flies and nematodes) and Lophotrochozoa (animals with a feeding structure called the lophophore, including snails and earthworms). We call this tree the L–E–D hypothesis (Fig. 1). Importantly, in this new tree, the coelom must have arisen more than once, or have been lost from some phyla.

Molecular analyses have been divided in

**Figure 1 Animals on trees: the two main hypotheses of the relationships between animal phyla.** a, The Acoelomata–Pseudocoelomata–Coelomata (A–P–C) phylogeny is supported by whole-genome studies, although complete genomes are available for only three animal phyla. In this scheme, flies (Arthropoda) and humans (Vertebrata) are more closely related to each other as members of the Coelomata than either is to nematodes (Pseudocoelomata)<sup>5,6</sup>. Based on morphology (there is no genome sequence), the Acoelomata are presumed to have separated from other animals before the divergence of the Pseudocoelomata and Coelomata. b, The new phylogeny, Lophotrochozoa–Ecdysozoa–Deuterostomia (L–E–D)<sup>1–3,7</sup>: using expressed-sequence-tag data, Philippe *et al.*<sup>7</sup> were able to include 12 animal phyla. Here, flies and nematodes are both members of the protostome group Ecdysozoa, distinct from the deuterostome humans.





their support for these competing hypotheses. Trees built using single genes from many species tend to support L–E–D<sup>8</sup>, but analyses using many genes from a few complete genomes support A–P–C<sup>5,6</sup>. The number of species represented in a phylogenetic study can have two effects on tree reconstruction. First, without genomes to represent most animal phyla, genome-based trees provide no information on the placement of the missing taxonomic groups. Current genome studies do not include any members of the Lophotrochozoa. More notably, if a species' genome is evolving rapidly, tree reconstruction programs can be misled by a phenomenon known as long-branch attraction<sup>9</sup>.

In long-branch attraction, independent but convergent changes (homoplasies) on long branches are misconstrued as 'shared derived' changes, causing artefactual clustering of species with long branches. Because these artefacts are systematic, confidence in them grows as more data are included, and thus genome-scale analyses are especially sensitive to long-branch attraction. Long branches can arise in two ways. One is when a distantly related organism is used as an 'outgroup' to root the tree of the organisms of interest. The other is when one organism of interest has a very different, accelerated pattern of evolution compared with the rest. Unfortunately for whole-genome studies, the usual outgroup, yeast, is very distantly related to animals, and *C. elegans* is a long-branch species<sup>5</sup>. Long-branch attraction will therefore tend to result in nematodes moving to the base of the tree, generating erroneous support for A–P–C. Not all whole-genome studies are tainted: analysis of rare insertions and deletions of genomic features (introns) in some animal genomes, characters that may be immune to the insidious charms of long-branch attraction, does not support A–P–C<sup>10</sup>.

Philippe *et al.*<sup>7</sup> have overcome these problems by using data from 'expressed sequence tags' (ESTs) in addition to complete genome sequences. Sequencing ESTs efficiently samples just the genes in any genome, avoiding the non-coding parts. The vastly lower cost of an EST project compared with sequencing a complete genome means that large numbers of ESTs have been generated for a much wider range of organisms, and we and others have been decorating the animal tree with EST data, including data from the neglected Lophotrochozoa<sup>11</sup>.

Using this expanded data set, Philippe *et al.*<sup>7</sup> find convincingly in favour of L–E–D (Fig. 1). They include many more data than previous non-genomic studies (35,371 amino acids from 146 genes) and more species than genome studies (35 species representing 12 animal phyla and 14 outgroups including choanoflagellates, thought to be the protozoan phylum most closely related to animals).

When only a distant outgroup (yeast) was

used<sup>7</sup>, nematodes emerge at the base of the tree. But with closer outgroups (protozoans related to animals, and jellyfish), nematodes cluster with arthropods, as predicted by the L–E–D hypothesis. In the complete data set, however, lophotrochozoan flatworms cluster with the ecdysozoan nematodes, and not with their supposed lophotrochozoan relatives (the molluscs and annelid worms). Suspecting that this was another long-branch artefact, Philippe *et al.* selectively eliminated genes expected to contribute most to long-branch attraction — those with a greater evolutionary rate in some species (such as nematodes) compared with others (such as deuterostomes). Indeed, as the most biased genes were removed, support for Ecdysozoa and Lophotrochozoa increased.

Will this be the last, defining statement in the controversy? There remain some unresolved problems with Philippe and colleagues' analysis, such as the position of the phylum Tardigrada (water bears). Tardigrades are unquestionably close to arthropods (they have eight stumpy legs), but appear as the sister phylum to the Nematoda. Have the nematodes lost the legs they once had, or are tardigrades misplaced?

Additionally, only 12 of the 35 animal phyla are currently represented: will addition of more phyla — particularly lophotrochozoan phyla — change the tree significantly? Have coeloms in protostome and deuterostome animals very different developmental

origins: have they arisen independently? Some really obscure bilaterally symmetrical animal phyla, such as acoel flatworms, are thought to have separated from the main animal lineage before the divergence of protostomes and deuterostomes. Will these illuminate the evolution of our own complex bodies?

Genome sequences from many other animals are now being gathered, and EST projects are under way or planned for many more. No doubt the tree will sprout new shoots — and new controversies — very soon.

Martin Jones and Mark Blaxter are at the Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Ashworth Laboratories, King's Buildings, Edinburgh EH9 3JT, UK.

e-mail: mark.blaxter@ed.ac.uk

1. Aguinaldo, A. M. A. *et al.* *Nature* **387**, 489–493 (1997).
2. Winnepenninckx, B. *et al.* *Mol. Biol. Evol.* **12**, 1132–1137 (1995).
3. Adoutte, A., Balavoine, G., Lartillot, N. & de Rosa, R. *Trends Genet.* **15**, 104–108 (1999).
4. Mushegian, A. R., Garey, J. R., Martin, J. & Liu, L. X. *Genome Res.* **8**, 590–598 (1998).
5. Blair, J. E., Ikeo, K., Gojobori, T. & Hedges, S. B. *BMC Evol. Biol.* **2**, 7 (2002).
6. Wolf, Y. I., Rogozin, I. B. & Koonin, E. V. *Genome Res.* **14**, 29–36 (2004).
7. Philippe, H., Lartillot, N. & Brinkmann, H. *Mol. Biol. Evol.* **22**, 1246–1253 (2005).
8. Peterson, K. J. & Eernisse, D. J. *Evol. Dev.* **3**, 170–205 (2001).
9. Felsenstein, J. *Syst. Zool.* **27**, 401–410 (1978).
10. Roy, S. W. & Gilbert, W. *Proc. Natl Acad. Sci. USA* **102**, 4403–4408 (2005).
11. Davison, A. & Blaxter, M. L. *Parasitology* **130**, 1–14 (2004).

## Technology

# Warm fusion

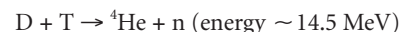
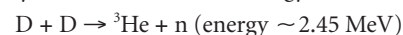
Michael J. Saltmarsh

A device that could fit in your lab-coat pocket uses nuclear fusion, and just a little heat, to produce neutrons. The advantages in simplicity and portability over conventional neutron generators could be considerable.

On page 1115 of this issue, Naranjo, Gimzewski and Putterman<sup>1</sup> report the successful demonstration of an intriguingly simple neutron generator that produces neutrons possessing an energy of 2.5 mega-electronvolts (MeV) from reactions involving the fusion of two nuclei of deuterium. This device, it must be stressed, will not generate net energy, and is not related to past controversies about 'cold fusion'.

Neutrons can penetrate significant quantities of matter, and interact primarily with the nucleus rather than the electronic structure of an atom. As a result, portable neutron generators have found a wide range of applications, including well-logging for oil exploration, and the screening of baggage for airline security. Several commercial devices are available that use fusion reactions of deuterium (D) and tritium (T), whose nuclei contain one and two

neutrons respectively (ordinary hydrogen nuclei have none). The reactions generate helium and a single neutron that carries away most of the reaction energy:



These neutron generators rely either on an ion beam from a miniature accelerator producing reactions in a solid target loaded with deuterium and/or tritium, or on the electrostatic confinement of a D–D or D–T plasma. In both cases high-voltage power is required, and the apparatus is fairly complex.

The device reported by Naranjo *et al.*<sup>1</sup> falls into the solid-target category, only without much of the complexity. Indeed, in some ways it is remarkably low-tech — the only input is a few tens of volts, to bias an electron-suppression grid, and some gentle

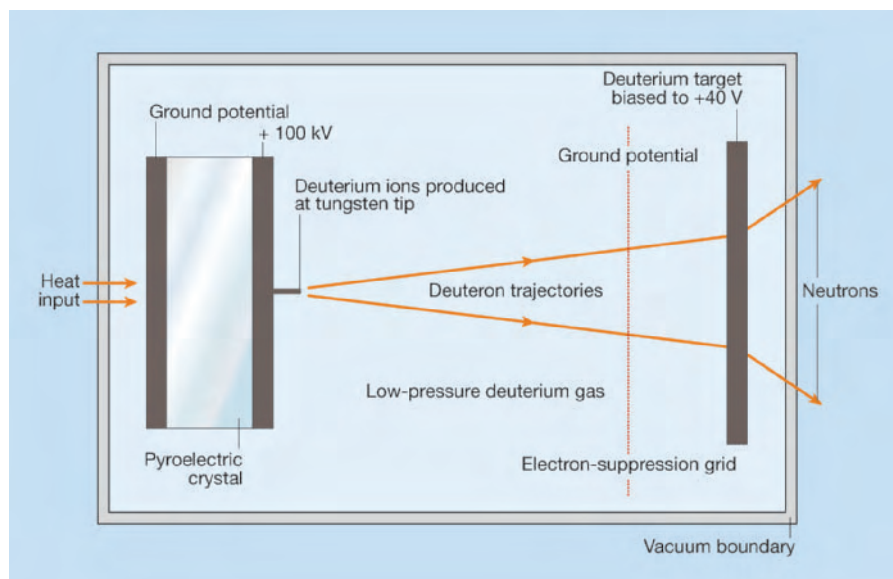


Figure 1 Naranjo and colleagues' apparatus for neutron generation<sup>1</sup>. The chamber is filled with deuterium gas at low pressure (0.7 pascals). As the crystal is heated, the potential builds across the crystal. Deuterium ions (deuterons) are generated at the tungsten tip, and accelerated towards the target; the electrons fall back to the crystal electrode. The ions strike the deuterium target ( $\text{ErD}_3$ ), and some generate 2.5-MeV neutrons. Electrons knocked from the target surface are repelled by the suppression grid and fall back on to the target rather than being accelerated back to the crystal.

heat (around 2 watts). A minute or two after the heat is applied, neutron emission starts, reaching a peak of about 1,000 per second; once the heat source is removed, the device gradually switches itself off. The key to the device's simplicity lies in the replacement of the miniature ion-source and accelerator in existing generators by a system based on a combination of two well-known phenomena — the pyroelectric effect and field ionization.

The pyroelectric effect — the fact that some materials become charged when heated — was probably first recorded in 314 BC by Theophrastus<sup>2</sup>, Aristotle's student and successor, from his studies of the gemstone tourmaline. More recently, various man-made materials have been investigated, and potentials of around 100,000 volts reported for crystals such as lithium tantalate ( $\text{LiTaO}_3$ ), with the emission of energetic electrons under suitable conditions. This effect was used by Brownridge<sup>3,4</sup> to produce a small pyroelectric X-ray generator, of which a commercial version, powered by a 9-volt battery, is now available<sup>5</sup>.

Field ionization of gases occurs when a potential difference of a few volts exists over atomic distances — equivalent to a field greater than 10 gigavolts ( $1 \times 10^{10}$  volts) per metre. The effect is widely used as the basis of field-ion microscopy. Modest voltages applied to electrodes of very small radius can produce these extremely high fields near the electrode tips, ensuring the ionization of essentially all gas molecules entering the high-field region.

Figure 1 shows how Naranjo *et al.* combined these effects to generate fusion

neutrons. The authors grounded one face of a 1-cm-thick pyroelectric crystal to the inside of a vacuum chamber containing deuterium gas at a pressure of 0.7 pascals (for comparison, Earth's atmospheric pressure is around  $10^5$  pascals). They then attached a tiny tungsten electrode to a plate on the positive face of the crystal. A solid target containing deuterium in the form of erbium deuteride ( $\text{ErD}_3$ ) was placed a few centimetres in front of this electrode.

Raising the temperature of the crystal at a rate of  $12.4^\circ\text{C}$  per minute changed the spontaneous polarization of the crystal, and raised the potential of the positive electrode at a rate of about 50 kilovolts per minute. As the potential rose, the field near the tungsten electrode increased to a value — around 25 gigavolts per metre — sufficient to produce field ionization of the deuterium gas. The positively charged ions (deuterium nuclei, or 'deuterons') produced in this process were accelerated towards the target across essentially the full potential generated by the crystal; the electrons stripped from the deuterium atoms by the ionization experienced a potential drop of only a few volts as they fell back to the crystal. On hitting the target on the opposite wall of the device, the energetic deuterons interacted with the deuterium target to produce 2.5-MeV neutrons via the  $\text{D} + \text{D}$  reaction.

The maximum current obtained in this experiment was about 4 nanoamperes, leading to a maximum neutron production rate of around 1,000 neutrons each second. The accelerating potential can be maintained only while the crystal temperature is changing; thus, the duration of the pulse



#### 100 YEARS AGO

The water-supply for the occupants of our huge prehistoric "camps" has always been somewhat of a mystery, and it has been suggested that they were only temporary refuges... But the watering of men and animals on the scale indicated by the areas enclosed would be a formidable task even for a day, and other explanations must be sought. The late General Pitt-Rivers, for example, held that the water-level of the combs was higher then than now, and streams would have been plentiful on the slopes; but, feeling the inadequacy of this view, he also had recourse to the dew-pond theory... An exposed position innocent of springs was selected, and straw or some other non-conductor of heat spread over the hollowed surface. This was next covered with a thick layer of well puddled clay, which was closely strewn with stones. The pond would gradually fill, and provide a constant supply of pure water, due to condensation during the night of the warm, moist air from the ground on the surface of the cold clay... Some ponds of this kind, no doubt of very early and perhaps of Neolithic date, may still be seen in working order.

From *Nature* 27 April 1905.

#### 50 YEARS AGO

In 1949 Burnet and Fenner postulated that antibody production against a particular antigen can be specifically suppressed by exposure to the same antigen during embryonic life... The decisive step which brought the principle of immunological tolerance from the sphere of Nature's eccentricity into the domain of an experimental method of possibly very wide applicability was the artificial production of a similar type of tolerance... The present investigations were carried out on birds by the method devised by Billingham, Brent and Medawar, and had a threefold purpose: (1) investigation of whether tolerance could be acquired to cells of foreign species; (2) if so, whether the tolerance would also extend to a virus from the donor animal to which the recipient was not normally susceptible; (3) whether a sexual cross between species which is not normally possible could be made so by means of acquired tolerance. Positive answers to the first two questions have been obtained; the answer to the third awaits the sexual maturation of the treated birds.

Morten Simonsen

From *Nature* 30 April 1955.

at this current level was limited to a few minutes by the attainable temperature rise. Although this output is too small for most applications, the authors outline plans to increase the yield to a million neutrons per second, comparable to that of some commercial portable neutron generators. Nevertheless, even at the level already attained, there are laboratory uses, such as measuring neutron detector response or for student practical demonstrations, for which a simple,

inexpensive, monoenergetic neutron source would be most valuable. ■

Michael J. Saltmarsh is mostly retired from Oak Ridge National Laboratory, PO Box 2008, Oak Ridge, Tennessee 37831-6368, USA. e-mail: msaltmarsh@aol.com

1. Naranjo, B., Gimzewski, J. K. & Putterman, S. *Nature* **434**, 1115–1117 (2005).
2. Donnay, G. *Acta Crystallogr. A* **33**, 927–932 (1977).
3. Brownridge, J. D. *Nature* **358**, 287–288 (1992).
4. Brownridge, J. D. & Raboy, S. *J. Appl. Phys.* **86**, 640–647 (1999).
5. www.amptek.com/coolx.html

## HIV

# Viral blitzkrieg

R. Paul Johnson and Amitinder Kaur

It takes years for AIDS to develop from the damage inflicted on the immune system by HIV or its simian counterpart. Surprisingly, as many as half of the body's memory T cells may die at a very early stage of infection.

**H**IV and the related simian immunodeficiency virus (SIV) cause AIDS by infecting the master regulatory cells of the immune system — T helper cells, better known as CD4<sup>+</sup> T cells. It is generally years before enough damage is done to this cellular army for full-blown AIDS to develop. Nevertheless, two reports in this issue<sup>1,2</sup> (pages 1093 and 1148) suggest that the outcome of the battle between SIV and its host may be determined by a dramatic opening salvo, in which the virus eliminates around half of the host's memory CD4<sup>+</sup> T cells within four days — thus setting the stage for a lengthy war of attrition.

CD4<sup>+</sup> T lymphocytes are so called because they express a receptor protein termed CD4; this is necessary for T-cell function, but has also been co-opted by HIV and SIV to gain entry to the cells. Previous results suggested that HIV/SIV replication is restricted

to a relatively small fraction (0.01–1%) of CD4<sup>+</sup> T cells in the chronic stages of infection<sup>3</sup>. This low frequency of infection seemed to reflect the fact that the viruses require the presence of a co-receptor in addition to CD4 to gain entry, and also that they replicate best in activated memory CD4<sup>+</sup> T cells (memory cells being those previously stimulated by foreign antigen). The preferred co-receptor for most HIV and SIV strains is CCR5, which is expressed only in a subset of memory CD4<sup>+</sup> T cells.

Despite the apparent low frequency of T-cell infection in chronic infection, seminal studies in SIV-infected monkeys<sup>4</sup> — subsequently confirmed in HIV-infected humans<sup>5,6</sup> — revealed a rapid and widespread depletion of CD4<sup>+</sup> T cells in the gut (mucosal T cells) during the first few weeks of infection. T cells in the blood or lymph nodes did not show the same degree of

depletion. This predilection of HIV and SIV for replicating in gut lymphocytes was felt to be a consequence of the relatively large populations of activated CD4<sup>+</sup> T cells at this site that express CCR5. However, the proportion of cells actually infected with SIV was not known; nor was it clear whether T-cell activation was in fact required for infection, or how T cells were killed.

Mattapallil *et al.*<sup>1</sup> have now examined the role of SIV in the depletion of CD4<sup>+</sup> T cells during early (acute) stages of infection. As previously reported<sup>4</sup>, SIV rapidly depleted CD4<sup>+</sup> T cells in the gut. Remarkably, however, 60–80% of memory CD4<sup>+</sup> T cells were concurrently depleted at all sites. Using a technique that can detect a single copy of SIV DNA, the authors determined that 30–60% of all memory CD4<sup>+</sup> T lymphocytes were infected with SIV within 10 days of infection, regardless of their location, and that most of these cells had disappeared 4 days later. These percentages far exceed the number of CD4<sup>+</sup> T cells that express CCR5, but the authors propose that this apparent contradiction may be resolved by their finding of low levels of CCR5-encoding messenger RNA in memory cells in which CCR5 protein could not be detected by flow cytometry. This implies that such cells may in fact express sufficient levels of CCR5 protein to render them permissive for SIV infection. Alternatively, SIV may be entering the cells using other co-receptors.

Li and colleagues' paper<sup>2</sup> provides a complementary perspective on this viral blitzkrieg. By identifying cells expressing SIV RNA in tissue sections, these investigators characterized the activation state of virus-producing cells (viral production requires SIV DNA to be transcribed into RNA). Consistent with their earlier work<sup>7</sup>, they found that most infected cells did not express markers of activation (CD25 or CD69), nor did they express Ki67 — a molecule found in

## Behavioural ecology

# Cue for kin

If you yourself can't breed, you can at least help your relatives with their offspring. Such altruistic behaviour occurs in long-tailed tits (*Aegithalos caudatus*, pictured), which Stuart Sharp and colleagues have studied to find out what cues enable a 'helper' to recognize kin. Their report appears elsewhere in this issue (*Nature* **434**, 1127–1130; 2005).

Adult long-tailed tits pair off and attempt to breed each year, but many don't succeed because of high rates of predation on the eggs or nestlings. The childless parents may then turn to assisting kin in feeding their brood — which makes sense in evolutionary terms but requires some

form of recognition system. Long-tailed tits are not the greatest of vocalists. They sing infrequently but do have an individually characteristic contact call, the 'churr', which develops even before fledging and is retained in the adult bird.

The first part of Sharp and colleagues' research involved the playback to individuals of churr calls belonging to a close relative and a non-relative, and a further two trials in which the frequency of these calls had been tweaked. The responses of birds to the untweaked calls of relatives differed significantly from their responses to the other three calls. From this, the authors

conclude that the churr call provides cues involved in kin recognition.

The most innovative part of the study, however, was an investigation into how much churr acquisition owes to nurture (learning) and nature (genetics). This took the form of swapping young birds between nests, so that adult birds were raising foster nestlings along with their true offspring. The churr calls of the fostered birds were the same as those of their nestmates, and unlike those of their biological siblings raised elsewhere — so it seems that the churr is in large part learned.

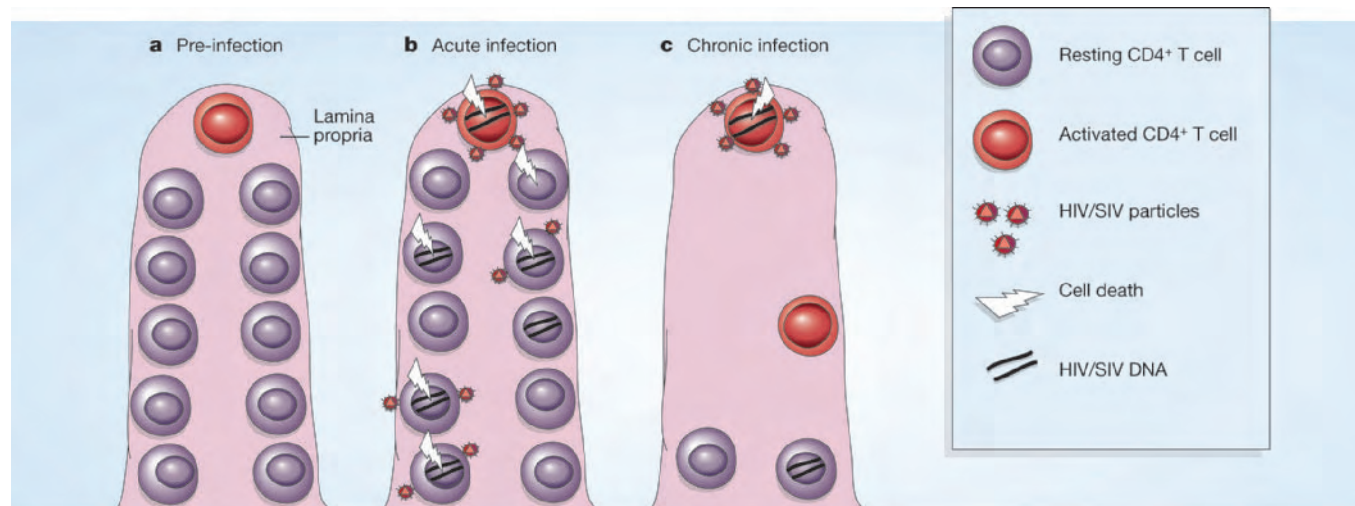
The pattern of helping observed in long-tailed tits is consistent with



the use of this learned cue in the great majority of cases. But this kin-recognition system evidently isn't flawless: in 6% of cases, an adult helped unrelated nestlings. As the authors point out, recognition systems are rarely perfect. **Tim Lincoln**

ANDREW MACCOLL





**Figure 1** Model for the depletion of gut CD4-expressing T lymphocytes by SIV. Most of the body's memory CD4<sup>+</sup> T helper cells are found in the gut, mainly in a compartment known as the lamina propria. **a**, Before infection, non-dividing (resting) CD4<sup>+</sup> T cells in the lamina propria outnumber activated T cells by 70-fold or more. **b**, In the initial (acute) phase of infection, the CD4<sup>+</sup> T-cell population in the gut is eliminated rapidly; the new papers reveal that up to 60% of these T cells are infected with SIV<sup>1</sup> and that most of the infected cells are resting rather than

activated<sup>2</sup>. Only a subset of infected cells that contain SIV DNA will express SIV RNA and produce viral particles. Such 'productively infected' cells, cells that contain viral DNA but not RNA, and uninfected cells may all be killed as a result of SIV infection, although the relative numbers of each are not known. **c**, During chronic infection, in response to the depletion of resting cells, the number of activated cells increases slightly. These cells now represent the dominant site of viral replication.

cells that have recently divided. Hence, these cells were deemed to be 'resting'.

Although these resting cells produce less virus than do Ki67-expressing cells, they vastly outnumber their activated counterparts and so serve as the major viral reservoir during this first phase of infection. Because of its ability to replicate in non-dividing cells, SIV can broaden its substrate base significantly, and deplete most CCR5<sup>+</sup> CD4<sup>+</sup> T cells in all lymphocyte compartments during acute infection. Once the resting memory-cell population is nearly eliminated, the activated population increases slightly, but now makes a larger contribution to total virus production (Fig. 1).

The idea that SIV (and presumably HIV) can infect resting lymphocytes seems to contradict the conventional wisdom that these viruses replicate primarily in activated cells. But mucosal lymphocytes (the main target cells) are probably better described as 'recently activated' rather than truly resting. Many such cells appear to be derived from blood cells that have recently divided and then migrated into the mucosa, where they lose Ki67 expression<sup>8</sup>. The dichotomous nomenclature of 'resting' and 'activated' T cells may obscure our ability to understand the range of cellular states that affects SIV/HIV replication. Further research is needed to determine the roles of viral proteins, host cytokine proteins, nucleotide pools and cellular resistance factors<sup>9,10</sup> in regulating viral replication in non-dividing lymphocytes.

Although there are many points of agreement between the new papers<sup>1,2</sup>, they differ on a key issue: the relative roles of direct versus indirect killing (death of virus-infected

versus uninfected cells). Mattapallil and colleagues' finding<sup>1</sup> of SIV DNA in up to 60% of memory CD4<sup>+</sup> T cells at peak suggests that nearly all T-cell death at early stages can be attributed to direct infection. But Li *et al.*<sup>2</sup> found a significantly lower peak frequency of cells expressing SIV RNA, and suggest that indirect mechanisms (mediated by the Fas–FasL cell-suicide pathway) also contribute to T-cell death in acute infection.

Differences in the sensitivity of the techniques used by the two groups to detect infected cells may contribute to this discrepancy. But another, perhaps more likely, possibility is that not all of the cells that contained SIV DNA were producing new viruses (indicated by the presence of SIV RNA). A similar situation occurs in chronic HIV infection, where fewer than 1 in 100 infected cells is 'productively infected'<sup>3</sup>.

Do these acute events determine the chronic course of infection? Owing to their focus on acute infection, neither study<sup>1,2</sup> allowed early events to be correlated with disease progression. So it is not known whether differences in the proportion of SIV-infected CD4<sup>+</sup> T cells — or in the size of the memory T-cell pool destroyed during peak replication — affect the extent of virus production during chronic infection, a key predictor of disease progression. The contribution of immune responses to the depletion of infected CD4<sup>+</sup> cells and the initial drop in peak viral load also remains controversial.

However, once acute viral replication has subsided, different viral and T-cell dynamics will come into play. Overall levels of virus production will fall, even though the shift in viral replication from resting to activated

T cells means that more viruses are produced per cell. The smaller size of the infected T-cell pool and the generation of virus-specific immune responses are both likely to contribute to the overall decrease in virus production. The body's ability to regenerate mucosal lymphocyte pools is also likely to be important in forestalling the onset of AIDS<sup>8</sup>.

As for vaccines, the findings reinforce the rationale for blocking HIV before it gains access to the fertile grounds of mucosal lymphocytes — and for reducing the initial burst of viral replication. Recent vaccine efforts have been aimed at activating other T cells, those expressing CD8, which are most effective in decreasing chronic levels of viral replication and slowing down disease progression<sup>11</sup>. But it may be more effective to stimulate the body to produce neutralizing antibodies that prevent the initial burst of replication — although this remains a challenge.

R. Paul Johnson and Amitinder Kaur are in the Division of Immunology, New England Primate Research Center, Harvard Medical School, Southborough, Massachusetts 01772, USA.

R. Paul Johnson is also at the Partners AIDS Research Center and Infectious Disease Unit, Massachusetts General Hospital, Boston. e-mail: paul\_johnson@hms.harvard.edu

1. Mattapallil, J. J. *et al.* *Nature* **434**, 1093–1097 (2005).
2. Li, Q. *et al.* *Nature* **434**, 1148–1152 (2005).
3. Haase, A. T. *Annu. Rev. Immunol.* **17**, 625–656 (1999).
4. Veazey, R. S. *et al.* *Science* **280**, 427–431 (1998).
5. Brenchley, J. M. *et al.* *J. Exp. Med.* **200**, 749–759 (2004).
6. Mehandru, S. *et al.* *J. Exp. Med.* **200**, 761–770 (2004).
7. Zhang, Z. *et al.* *Science* **286**, 1353–1357 (1999).
8. Picker, L. J. *et al.* *J. Exp. Med.* **200**, 1299–1314 (2004).
9. Unutmaz, D., KewalRamano, V. N., Marmon, S. & Littman, D. R. *J. Exp. Med.* **189**, 1735–1746 (1999).
10. Chiu, Y.-L. *et al.* *Nature* doi:10.1038/nature03493 (2005).
11. Robinson, H. L. *Nature Rev. Immunol.* **2**, 239–250 (2002).

## Obituary

## Maurice R. Hilleman (1919–2005)

During the twentieth century, life expectancy in industrialized countries almost doubled. Improvements in nutrition and sanitation certainly played a major role, but the central factor responsible was the development of childhood vaccines. A particular name can be attached to this achievement — that of Maurice R. Hilleman.

Hilleman ranks with the most brilliant and productive creators of vaccines of all time, with a stature comparable to that of Edward Jenner and Louis Pasteur. But his career was hardly foreseeable. Born in remote Miles City, Montana, as the last of eight children (his mother and twin sister dying during the birth), Hilleman grew up on his uncle's farm, "where things got done" — as Hilleman himself often said. He was later renowned for bringing the same attitude to work in the laboratory. He would never go home before an experiment was finished, and whenever possible was continually driven to translate results in basic or applied research to practical ends.

Growing up on a lonely ranch, Hilleman soon became interested in biology. He talked to the farm animals, trying to hypnotize and train them, but eventually discovered Charles Darwin's *Origin of Species*. Caught reading this book in church did nothing to enhance his popularity in a community of staunch Lutheran Protestants.

Indeed, Hilleman's career almost never got started. When he graduated in 1937, his family had no plans for him to go to college and instead he took a job as a stock boy at a local store. However, one of his older brothers, returning home for the summer, persuaded Hilleman that his talents were being wasted and he enrolled at Montana State University on a scholarship. There, in order to be finished by mid-term, he would often run a semester's worth of organic-chemistry experiments simultaneously. He won a fellowship to the University of Chicago and finished in 1944 with a prizewinning PhD dissertation on the bacterium *Chlamydia*. At that time, it was not possible to accurately subtype bacterial or viral families, but Hilleman developed the appropriate techniques by immunizing animals with different isolates and studying antigenic cross-recognition by the resulting antibodies.

Hilleman's scientific successes continued when he joined the pharmaceutical company E. R. Squibb in 1944. Because of the danger of Japanese encephalitis in the Pacific offensive during



### Creator of vaccines that continue to prevent death and disease on a global scale

the Second World War, his first challenge was to develop a vaccine against this viral disease. He completed the task within a few months, then cooperated with Nobel laureate Wendell Stanley from The Rockefeller University to develop the first highly immunogenic influenza vaccine with low toxicity.

In 1948, Hilleman left industry to concentrate on basic research at the Walter Reed Army Institute of Research in Washington. Among many other achievements, he discovered antigenic drift in influenza viruses, the phenomenon by which point mutations allow the virus to evade the full force of the continuously evolving herd immunity. He also discovered antigenic shift, that rare genetic reassortment that can occur during influenza double-infections and that leads to influenza strains against which herd immunity does not yet exist.

In 1957, this knowledge allowed Hilleman to warn the US authorities and the World Health Organization (WHO) that an influenza pandemic was about to emerge, a threat that had been overlooked by the WHO. He had noticed a report in *The New York Times* of a severe outbreak of respiratory disease in Hong Kong. Hilleman obtained samples, isolated a new influenza

strain and convinced manufacturers to produce 40 million doses of a vaccine. When, as predicted by Hilleman, the new virus reached the United States in September 1957, those at greatest risk were already being immunized.

Hilleman's productivity really shifted into high gear after 1957, when he joined Merck Research Laboratories in West Point, Pennsylvania. He was given full responsibility to plan and direct vaccine research and development. During the next 30 years, Hilleman and his co-workers produced more than 40 viral and bacterial vaccines. They developed the concept of combined live virus vaccines, which culminated in the global marketing of the measles–mumps–rubella (MMR) vaccine that is estimated to save the lives of almost two million children each year. Other remarkable vaccines against hepatitis A and hepatitis B followed. His hepatitis B vaccine was also the first anticancer vaccine, as it prevents hepatocellular carcinoma, a consequence of infection.

The following episode exemplifies Hilleman's dedication to science. One night in 1963, he was awakened by his little daughter Jeryl Lynn who was sick with a sore throat. He rightly diagnosed mumps, took a throat swab and isolated the virus. This virus strain was then used to develop a vaccine, and Hilleman, assisted by his wife Lorraine, a trained nurse, immunized their second daughter Kirsten in one of the first successful clinical trials.

Hilleman's accomplishments in basic and applied research have been of seminal importance to medical biology and human health. They have increasingly received recognition for their insight and ingenuity — it is no exaggeration to state that they have changed the world. Together with antibiotics such as penicillin, the products of Hilleman's hard work and skill have prevented and continue to prevent the suffering of billions, and the deaths of millions. In acknowledgement of his work, he received many honours and prizes, among them the Albert Lasker Public Service Award and the US National Medal of Science.

Hilleman was an indefatigable worker, still publishing and giving seminars until a few weeks before his death on 11 April. He was our godfather in the scientific community: usually quiet, yet precisely intervening and creative in discussions, approachable for youngsters and always helpful, he was our benchmark. His many friends will remember him, as should the whole of mankind.

Reinhard Kurth

Reinhard Kurth is at the Robert-Koch-Institut, Nordufer 20, 13353 Berlin, Germany.  
e-mail: schwarzk@rki.de

## Colloid science

### Double bubbles are no trouble

*Science* **308**, 537–541 (2005)

Emulsions — droplets of one liquid suspended in another — are widely used in technology, food processing and cosmetics. Double emulsions, in which each droplet encapsulates a smaller droplet of another liquid, provide further possibilities. For example, they allow the release of drugs or flavourings from the inner droplet to be governed by the size and composition of the outer droplet. Gelling or polymerizing the outer droplet results in the formation of a robust capsule for controlled release.

It has been hard to control the absolute and relative sizes of both droplets in such double emulsions. But A. S. Utada *et al.* have now come up with an answer: a versatile microfluidic system that can deliver not only precisely determined droplet sizes, but also specified numbers of inner droplets in each outer one.

The inner, outer and surrounding fluids are all squirted simultaneously and coaxially into the neck of a tapered capillary, with the innermost fluid being delivered into the mix through a separate micropipette. A theoretical model allows the sizes of the droplets to be predicted from the instrumental geometry and liquid flow rates, providing anything from thin-walled fluid capsules to multi-compartment polymer vesicles.

Philip Ball

## Neurobiology

### Astrocytes star in inhibition

*J. Neurosci.* **25**, 3638–3650 (2005)

Astrocytes are well known for their structural role in the nervous system: these star-shaped cells hold neurons in place. They also contribute to neuronal function, including the activity of synapses — junctions between neurons that allow them to communicate with each other. Over the past decade, the role of astrocytes in the development of excitatory synapses has been revealed. It now seems that they modulate inhibitory synapses as well.

Sarina B. Elmiah *et al.* investigated the influence of astrocytes on synapses that use  $\gamma$ -aminobutyric acid (GABA) — the principal inhibitory neurotransmitter in the human brain. Drugs that target GABA receptors include anaesthetics and anti-epileptics, hence the additional significance in understanding how these synapses develop.

The authors cultured astrocytes with neurons from the rat hippocampus. They found that, in the absence of astrocytes, inhibitory synaptic terminals were rare by the fourth day *in vitro*. Conversely, their presence led to a sevenfold increase in the

number of these terminals at this same stage. Astrocytes also produced an increase in the number and synaptic localization of GABA-receptor clusters. Further experiments led the authors to suggest that astrocytes may achieve these effects in part by indirectly stimulating signalling between neurons that is mediated by neurotrophin proteins.

Roxanne Khamisi

## Materials science

### Carbon through the phases

*Phys. Rev. Lett.* **94**, 145701 (2005)

What happens to carbon when the heat is turned up and the pressure on? Carbon is subjected to extreme conditions in many geophysical and industrial processes: using a semi-empirical model known as

the ‘long-range carbon bond-order potential’, Luca M. Ghiringhelli *et al.* have looked into carbon’s behaviour in such situations.

Two very different ‘phases’ of pure carbon are familiar: graphite (the ‘lead’ in pencils) is usefully soft and opaque, whereas the less-everyday diamond is famously hard and clear. Squeezing graphite under high pressure eventually turns it into diamond, and diamond decays to graphite at room temperature and pressure — although extremely slowly.

Ghiringhelli and colleagues use their model to compute the temperatures and pressures at which transitions between graphite, diamond and a third — liquid — phase of carbon occur in conditions up to 12,000 kelvin and 400 gigapascals (4 million times Earth’s atmospheric pressure). Experimental data under extreme conditions are scarce, but where they do exist their agreement with the model is good. The results of the simulation also allow the hotly disputed existence of two distinct liquid phases of carbon to be discounted, the authors claim.

Andreas Tröbsinger

## Molecular biology

### Quick start

*Mol. Cell* **18**, 171–183 (2005)

At least 2,500 genes are switched on in just six minutes when cells of budding yeast leave the resting state to begin the cell division cycle, according to Marijana Radonjic, Jean-Christophe Andrau and colleagues.

Eukaryotic cells, which include yeast and mammalian cells, spend most of their lives in the resting state, with their ability to proliferate maintained but inhibited. How this quiescent condition is regulated has implications for understanding cancer, development and ageing.

From Radonjic and colleagues’ survey of yeast gene activity, it seems that RNA polymerase II, the enzyme responsible for much gene transcription, is poised and raring to go while the cells are resting. This is contrary to models of gene activation in which transcription of most genes is limited by the rate at which RNA polymerase II is recruited to the start site of the gene.

Radonjic *et al.* found that, in quiescent yeast cells, the enzyme was already predominantly bound next to the start sites of hundreds of genes that were immediately switched on as the cells began division. The authors suggest that such positioning allows a cell to respond rapidly, and begin growing, when it encounters nutrients in an environment where organisms compete strongly for resources.

Helen Dell

## Virology

### Polar flu

*Biol. Lett.* doi:10.1098/rsbl.2004.0253 (2005)

Newly spawned particles of influenza virus are released from only one face of virus-infected cells (pictured). Debra Elton *et al.* propose that this polar release has its roots in the asymmetric distribution of a viral protein in the cell nucleus.

The flu virus infects the epithelial cells that line the respiratory tract. It then coerces the cells into making new viral particles, which are released only from the exposed cell surface (the apical surface). Virus production begins with the

IMAGE  
UNAVAILABLE  
FOR COPYRIGHT  
REASONS

replication of the viral genome in the nucleus, giving numerous RNA molecules. These are then packaged with the viral nucleoprotein (NP) and exported from the nucleus.

Using infected cells in culture, together with staining techniques, Elton *et al.* found that when export was inhibited, NP accumulated at the nuclear periphery in an apical distribution. The same occurred temporarily when export was not inhibited, or when NP alone was transfected into uninfected cells. The latter finding suggests that NP must be interacting with something in the cell that is similarly distributed. One potential candidate, the authors speculate, is chromatin — the protein-bound complex into which nuclear DNA is packaged.

Amanda Tromans



# High-speed integrated nanowire circuits

Inexpensive sophisticated circuitry can be 'painted' on to plastic or glass substrates.

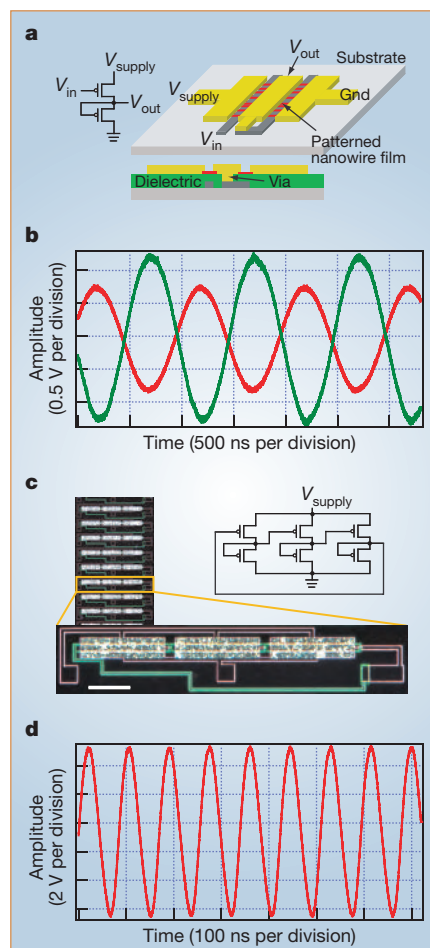
Macroelectronic circuits made on substrates of glass or plastic could one day make computing devices ubiquitous owing to their light weight, flexibility and low cost<sup>1</sup>. But these substrates deform at high temperatures so, until now, only semiconductors such as organics and amorphous silicon<sup>2</sup> could be used, leading to poor performance. Here we present the use of low-temperature processes to integrate high-performance multi-nanowire transistors into logical inverters and fast ring oscillators on glass substrates. As well as potentially enabling powerful electronics to permeate all aspects of modern life, this advance could find application in devices such as low-cost radio-frequency tags and fully integrated high-refresh-rate displays.

The mobility of single-crystal semiconducting nanowires is comparable to that of computer-grade silicon<sup>3</sup>. Multi-nanowire transistors<sup>4,5</sup> — analogous to thin-film transistors — can be assembled from solution on pieces of glass and plastic<sup>6</sup>. For many applications, however, fully interconnected nanowire devices that function as viable circuit elements operating at high frequencies will be required.

We integrated two nanowire thin-film transistors to generate inverters (Fig. 1a), which we made in a parallel process over glass substrates by using standard photolithography techniques (for methods, see supplementary information). The process gives a high yield of devices that show reliable, well defined signal inversion under direct current conditions (see supplementary information). Investigation of the alternating-current response of these inverters (Fig. 1b) shows that the gain, or signal amplification, is greater than unity, and the expected phase inversion is achieved when these devices are driven by a 1-MHz sine wave at a supply of 15 V. As signal propagation in an integrated system requires gain that is greater than unity, these results and the high reproducibility of our nanowire transistors suggest that fully interconnected nanowire oscillators could operate in the megahertz regime.

Our ring oscillators consist of three inverters in series (Fig. 1c), where the input of each inverter is connected to the output of the previous device, with a feedback loop to complete the ring. The necessary on-chip integration is achieved during fabrication and does not require any external wiring. We characterized the nanowire ring oscillators on glass substrates and found that the output-voltage oscillations were stable and self-sustained.

The devices show a maximal oscillation frequency of 11.7 MHz, corresponding to a



**Figure 1** Alternating-current properties of integrated multi-nanowire circuits on glass. **a**, Circuit diagram and schematics of the multi-nanowire inverters. Labelled voltages are bias ( $V_{supply}$ ), input ( $V_{in}$ ) and output ( $V_{out}$ ) voltage. The dielectric is omitted in the perspective schematic for clarity. 'Via' indicates one of a pattern of holes in the dielectric layer that are used to connect different metal layers. **b**, Output waveform (green) of an inverter fabricated on glass driven by a 1-MHz sine wave (red) with  $V_{supply} = 15$  V. **c**, Optical images and circuit diagram of nanowire ring oscillators. The gate level edge, source-drain level edge and nanowires appear green, pink and white, respectively, in dark field. Scale bar, 100  $\mu$ m. **d**, Oscillation of 11.7 MHz in a ring oscillator structure with  $V_{supply} = 43$  V.

stage delay of 14 ns (Fig. 1d). Significantly, all devices measured on glass have oscillation frequencies at or above 10 MHz. Furthermore, nanowire oscillators made on glass substrates have higher frequencies than devices made on silicon substrates (see supplementary information). This could be a key advantage for nanowire circuits as the properties of devices made with other materials often degrade upon transfer to non-crystalline substrates<sup>7</sup>.

The stable oscillation frequencies seen for our nanowire-based devices are many orders of magnitude larger than previous ring oscil-

lators based on nanoscale building blocks. For example, carbon-nanotube devices have oscillation frequencies from 5 to 220 Hz (refs 8, 9). Although this is not an intrinsic limit for nanotubes, it highlights how important reproducible material properties are for the successful creation of integrated, high-performance devices.

It is interesting to compare these nanowire-device features with those of organic ring oscillators, given that the active material in both can be deposited at ambient temperatures from liquid solutions. Reported stage-delay times for organic ring oscillators are typically longer than 300 ns (ref. 10) and therefore substantially ( $20\times$ ) slower than those we obtain for nanowires on glass. Comparable results have been seen for other semiconductors with low synthesis temperatures; to our knowledge, the fastest reported stage delay for amorphous silicon ring oscillators is 210 ns (ref. 11).

Our integrated nanowire-based transistors open the way to a variety of electronic applications; the techniques we describe are all compatible with low-deformation-temperature materials such as plastics, broadening the scope for design. One limitation is that supply voltages of 35 volts or more are required to achieve stable oscillations, but device structures could be improved by incorporating higher- $k$  dielectrics, more advanced nanowire materials and reduced channel lengths. This would enable them to be operated at lower voltages and much higher frequencies, taking low-cost electronics to high-performance computing levels.

**Robin S. Friedman\***, **Michael C. McAlpine\***, **David S. Ricketts†**, **Donhee Ham†**, **Charles M. Lieber\*†**

\*Department of Chemistry and Chemical Biology, †Division of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA  
e-mail: cml@cmliris.harvard.edu

- Gelinck, G. H. *et al.* *Nature Mater.* **3**, 106–110 (2004).
- Street, R. A. *Technology and Applications of Amorphous Silicon* (Springer, New York, 2000).
- Cui, Y., Zhong, Z., Wang, D., Wang, W. U. & Lieber, C. M. *Nano Lett.* **3**, 149–152 (2003).
- Jin, S. *et al.* *Nano Lett.* **4**, 915–919 (2004).
- Duan, X. *et al.* *Nature* **425**, 274–278 (2003).
- McAlpine, M. C. *et al.* *Nano Lett.* **3**, 1531–1535 (2003).
- Shimoda, T., Inoue, S. & Utsunomiya, S. *Proc. SPIE* **4295**, 52–59 (2001).
- Bachtold, A., Hadley, P., Nakanishi, T. & Dekker, C. *Science* **294**, 1317–1320 (2001).
- Javey, A., Wang, Q., Ural, A., Li, Y. & Dai, H. *Nano Lett.* **2**, 929–932 (2002).
- Clemens, W., Fix, W., Ficker, J., Knobloch, A. & Ullmann, A. *J. Mater. Res.* **19**, 1963–1973 (2004).
- Hiranaka, K., Yamaguchi, T. & Yanagisawa, S. *IEEE Electron Device Lett.* **7**, 224–225 (1984).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

## Earth science

# Microseismicity data forecast rupture area

On 28 September 2004 there was an earthquake of magnitude 6.0 at Parkfield, California. Here we show that the size distribution of the micro-earthquakes recorded in the decades before the main shock occurred allowed an accurate forecast of its eventual rupture area. Applying this approach to other well monitored faults should improve earthquake hazard assessment in future.

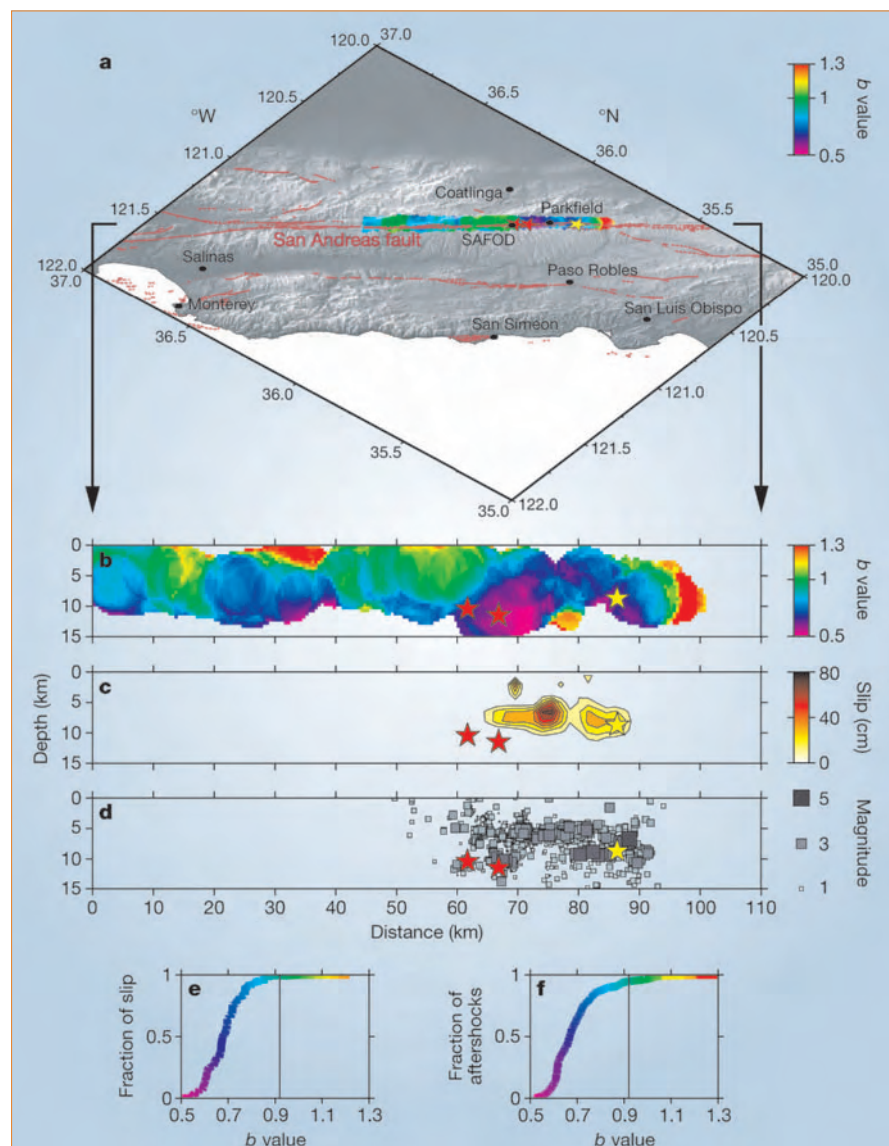
The Parkfield event fulfilled, in terms of location and size, the forecast made 20 years ago in the Parkfield earthquake-prediction experiment<sup>1</sup>. However, the course of the rupture came as a surprise: it started in the southern end of the rupture zone of the last magnitude-6.0 event of 1966 and continued mainly to the north. The timing of the earthquake was not predicted, and no short-term precursors were reported before the event.

We are able to demonstrate that the rupture area of the Parkfield event accurately matches our earlier forecast<sup>2</sup>, which was derived from the size distribution of the microseismicity in the three decades before the event.

The cumulative earthquake-size distribution is commonly described by a power law:  $\log N = a - bM$ , where  $N$  is the cumulative number of earthquakes of magnitude  $M$  or greater,  $a$  is the earthquake productivity of a volume, and  $b$  is the relative size distribution<sup>3</sup>. In the laboratory,  $b$  values have long been known to be inversely dependent on differential stress<sup>4</sup>. If this dependence holds in the Earth's crust, then measurements of spatial and temporal changes in  $b$  could act as a 'stressmeter' to help image asperities — the highly stressed patches in faults where future ruptures are likely. The Parkfield event in September 2004 allowed us to test this idea<sup>2,5</sup>.

Figure 1a is a map view of the fault that shows the area of low  $b$  values around Parkfield. Figure 1b is a high-resolution image of  $b$  values along the fault<sup>5</sup>, obtained from seismic activity in the magnitude range 1.3 to 5 between January 1981 and September 2004. We interpret the unusually low  $b$  values in the Parkfield area as an indication of highly stressed patches in the fault (Fig. 1b). The preliminary slip distribution of the 2004 main shock<sup>6</sup>, and the aftershocks that occurred in the first week, are shown in Fig. 1c, d. The areas of low  $b$  value contain 99% of the slip and 95% of the aftershocks (Fig. 1e, f). Areas of high  $b$  value, which in the past we interpreted as creeping sections, act as barriers against rupture propagation. Two magnitude-5 aftershocks occurred within two days of the main event in the last remaining area of low  $b$  value that had been left unruptured by the main shock (Fig. 1b, c).

The close match we find between the area



**Figure 1** Cross-sectional views of the San Andreas fault at Parkfield. **a**, Topographic map of the Parkfield region with mapped  $b$  values. Red lines mark faults. Yellow stars, magnitude-6.0 main shock; red stars, magnitude-5.0 aftershocks. **b**, Microseismicity  $b$  values from 1981 to 27 September 2004 mapped along the cross-section. **c**, Preliminary slip distribution for the main shock that occurred on 28 September 2004. **d**, Aftershocks (28 September to 5 October) of the magnitude-6.0 event on 28 September 2004; symbol size is proportional to magnitude. **e**, Cumulative slip distribution as a function of  $b$  value; line colour corresponds to the  $b$  value. The vertical line indicates the regional average  $b$  value of 0.92. **f**, As **e**, but showing the cumulative distribution of aftershocks.

of low  $b$  values and the observed rupture area indicates that  $b$  values may indeed act as stressmeters for the Earth's crust. Our observations support previous retrospective studies<sup>7</sup> that correlated patches of low  $b$  value, based on recent microseismicity, with the main shocks at Izmit (Turkey), Kanto (Japan), the San Jacinto–Elsinore fault zone (California) and Morgan Hill (California).

The combined evidence indicates that  $b$  values could be used for accurately predicting rupture areas: although the timing of earthquakes remains unpredictable, precise forecasting of the location and size of events is becoming a possibility.

**Danijel Schorlemmer, Stefan Wiemer**

Swiss Seismological Service, ETH Zürich, 8093 Zürich, Switzerland  
e-mail: danijel@sed.ethz.ch

- Roeloffs, E. & Langbein, J. *Rev. Geophys.* **32**, 315–335 (1994).
- Wiemer, S. & Wyss, M. *J. Geophys. Res.* **102**, 15115–15128 (1997).
- Gutenberg, B. & Richter, C. F. *Bull. Seismol. Soc. Am.* **34**, 185–188 (1944).
- Scholz, C. H. *Bull. Seismol. Soc. Am.* **58**, 399–415 (1968).
- Schorlemmer, D., Wiemer, S. & Wyss, M. *J. Geophys. Res.* **109**, B12307 (2004).
- Ji, C. *et al.* *EOS Trans. AGU* **85**(47), Fall Meet. Suppl., Abstr. S53D-04 (2004).
- Wiemer, S. & Wyss, M. *Adv. Geophys.* **45**, 259–302 (2002).

Competing financial interests: declared none.

## brieF communications arising online

www.nature.com/bca

## Fisheries: Decline of Pacific tuna populations exaggerated?

J. Hampton, J. R. Sibert, P. Kleiber, M. N. Maunder, S. J. Harley (doi:10.1038/nature03581)

**Reply:** R. A. Myers, B. Worm (doi:10.1038/nature03582)



Fisheries

# Decline of Pacific tuna populations exaggerated?

Arising from: R. A. Myers & B. Worm *Nature* **423**, 280–283 (2003)

**T**una have been the target of large-scale industrial fisheries in the Pacific Ocean and elsewhere since the 1950s. In their analysis of Japanese longline-fishery catch-per-unit-effort (CPUE) data, Myers and Worm<sup>1</sup> conclude that the community (species-aggregated) biomass of large pelagic fish, mainly tunas, was reduced by 80% during the first 15 years of exploitation and is now at 10% of pre-industrial levels. We show here that an assumption critical to this conclusion — namely, that Japanese longline CPUE acts as an accurate index of community biomass — is invalid. Our results indicate that biomass decline and fishing impacts are much less severe than is claimed by Myers and Worm<sup>1</sup>.

Interpretation of the species-aggregated CPUE as an index of community biomass rests on the assumption that catchability (a coefficient specifying the proportionality between CPUE and abundance) is constant across species and over time. The former is unrealistic because, among other things, the species have different depth distributions and hence different vulnerability to longline gear<sup>2</sup>. The evolution of tuna longline fisheries in all oceans<sup>3</sup> has seen changes in fishing strategies (and hence catchability) as different species have been targeted. In the early 1960s, Japanese longliners changed from targeting albacore (*Thunnus alalunga*) and yellowfin (*T. albacares*) for the canned-tuna market to bigeye (*T. obesus*) and yellowfin tuna for the Japanese sashimi market<sup>3</sup>. Japanese longline CPUE for albacore declined rapidly not because of declining albacore abundance, but because of this change in species targeting. By contrast, Taiwanese longliners have consistently targeted albacore in subequatorial waters of all oceans, and their CPUE provides a better index of albacore abundance. These results show that CPUE has declined by 50% over 40 years in the South Pacific, but they do not replicate the rapid and much larger decline in CPUE in the 1960s evident in the Japanese data (Fig. 1a).

The Myers and Worm analysis<sup>1</sup> excludes data from the equatorial Pacific, where the highest catches are taken and which is the core habitat for tropical tunas. When these data are included, yellowfin-tuna CPUE in the western Pacific is seen to decline by 70% over 50 years, during which time annual catches by longline and other methods increase from insignificant levels in the early 1950s to more than 400,000 tonnes by the late 1990s (Fig. 1b). By contrast, the CPUE for bigeye tuna has been stable for over 40 years, despite continuously increasing catch (Fig. 1c). Changes in fishing strategies

designed to target the deeper-swimming and higher-value bigeye tuna occurred during the 1970s (ref. 3), making it unlikely that CPUE accurately reflects changes in abundance for either species unless it is adjusted to account for the shift in targeting<sup>4</sup>. Unadjusted Japanese longline CPUE tends to overestimate abundance decline for yellowfin tuna and underestimate abundance decline for bigeye tuna.

Stock assessments rely on a range of data in addition to CPUE, including catch, size composition, tagging and biological data. When stock-assessment models<sup>5,6</sup> that consider all the available data are applied to Pacific tunas, fishery-induced declines in abundance during the 1950s and 1960s of the magnitude proposed by Myers and Worm<sup>1</sup> are found to be extremely unlikely<sup>7–12</sup>. Moreover, where declines do occur, they are not, as claimed by Myers and Worm, due exclusively to fishing. It is impossible, for example, under conventional population-dynamics theory to attribute the pre-1970 decline in yellowfin CPUE to fishing at a time when the total catches were less than one-tenth of today's catches. In summary, the trends in catches and CPUE (Fig. 1) and the results of stock-assessment modelling show that the basic assumption of Myers and Worm that CPUE is proportional to

abundance is incorrect and inconsistent with any known population-dynamics effect.

Fisheries management cannot be based solely on examination of CPUE trends. In our opinion, management advice for Pacific tunas based on the conclusions of Myers and Worm<sup>1</sup> would be misleading. For example, current stock assessments indicate that bigeye is probably overexploited<sup>8,12</sup>, that yellowfin is fully exploited<sup>7,11</sup> and that southern albacore is lightly exploited<sup>10</sup>; management priorities inferred from these assessments would be reversed if assessments were based only on examination of Japanese longline CPUE. Also, the implication that all CPUE and abundance decline is fishery induced ignores the impact of environmentally induced recruitment variation<sup>13–15</sup>. For fisheries management to be effective, it is critical to discriminate between the effects on pelagic fish populations of environmental factors and fishing.

**John Hampton\*, John R. Sibert†, Pierre Kleiber‡, Mark N. Maunder§, Shelton J. Harley§**

\*Oceanic Fisheries Programme, Secretariat of the Pacific Community, BP D5, 98848 Noumea, New Caledonia

e-mail: johnh@spc.int

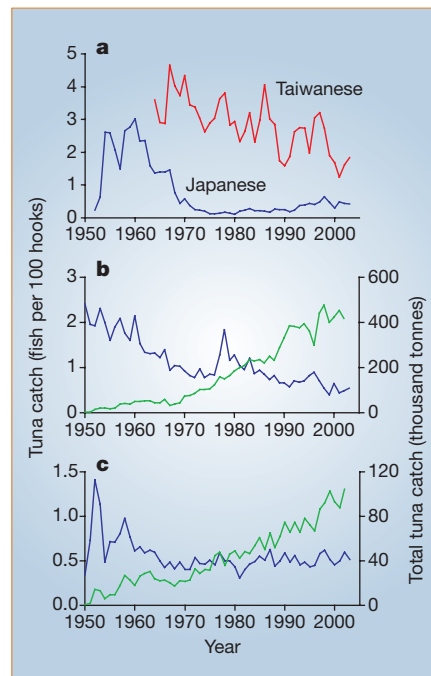
†Pelagic Fisheries Research Program, University of Hawaii, Honolulu, Hawaii 96822, USA

‡National Marine Fisheries Service, Pacific Islands Fisheries Science Center, Honolulu, Hawaii 96822-2396, USA

§Inter-American Tropical Tuna Commission, La Jolla, California 92037-1508, USA

doi:10.1038/nature03581

1. Myers, R. A. & Worm, B. *Nature* **423**, 280–283 (2003).
2. Brill, R. W. & Lutcavage, M. E. *Am. Fish. Soc. Symp.* **25**, 179–198 (2001).
3. Sakagawa, G. T., Coan, A. L. & Bartoo, N. W. *Mar. Fish. Rev.* **49**, 57–66 (1987).
4. Bigelow, K. A., Hampton, J. & Miyabe, N. *Fish. Oceanogr.* **11**, 143–155 (2002).
5. Hampton, J. & Fournier, D. *Mar. Freshwat. Res.* **52**, 937–963 (2001).
6. Maunder, M. N. & Watters, G. *Inter-Am. Trop. Tuna Comm. Bull.* **23**, 433–582 (2003).
7. Hampton, J. & Kleiber, P. *Stock Assessment of Yellowfin Tuna in the Western and Central Pacific Ocean* (16th Meeting of the Standing Committee on Tuna and Billfish, Secretariat of the Pacific Community, Noumea, New Caledonia, 2003). <http://www.spc.int/OceanFish/Html/SCTB/SCTB16/yft1.pdf>
8. Hampton, J., Kleiber, P., Takeuchi, Y., Kurota, H. & Maunder, M. N. *Stock Assessment of Bigeye Tuna in the Western and Central Pacific Ocean, with Comparisons to the Entire Pacific Ocean* (16th Meeting of the Standing Committee on Tuna and Billfish, Secretariat of the Pacific Community, Noumea, New Caledonia, 2003). <http://www.spc.int/OceanFish/Html/SCTB/SCTB16/bet1.pdf>
9. Kleiber, P., Hinton, M. & Uozumi, Y. *Mar. Freshwat. Res.* **54**, 349–360 (2003).
10. Labelle, M. & Hampton, J. *Stock Assessment of Albacore Tuna in the South Pacific Ocean* (16th Meeting of the Standing Committee on Tuna and Billfish, Secretariat of the Pacific Community, Noumea, New Caledonia, 2003). <http://www.spc.int/OceanFish/Html/SCTB/SCTB16/alb1.pdf>
11. Maunder, M. N. & Harley, S. J. *Inter-Am. Trop. Tuna Comm., Stock Assessment Report* **4**, 5–119 (2004). <http://www.iattc.org/StockAssessmentReportsENG.htm>
12. Harley, S. J. & Maunder, M. N. *Inter-Am. Trop. Tuna Comm., Stock Assessment Report* **4**, 120–286 (2004). <http://www.iattc.org/StockAssessmentReportsENG.htm>
13. Lehodey, P., Chai, F. & Hampton, J. *Fish. Oceanogr.* **12**, 483–494 (2003).



**Figure 1** Tuna catch per unit effort (CPUE) in the Pacific Ocean. **a**, Albacore tuna CPUE by Taiwanese (red) and Japanese (blue) longliners in the Pacific Ocean, south of the Equator. **b**, Yellowfin and **c**, bigeye tuna CPUE by Japanese longliners (blue) and catch by all fleets (green) in the western Pacific Ocean (west of 150° W, south of 20° N).



14. Ravier, C. & Fromentin, J.-M. *Fish. Oceanogr.* **13**, 145–160 (2004).

15. Chavez, F. P., Ryan, J., Lluch-Cota, S. E. & Niquen, C. *Science* **299**, 217–221 (2003).

**Myers and Worm reply** — Hampton and colleagues<sup>1</sup> challenge one aspect of our report on global declines in predatory fish communities<sup>2</sup>. They posit that Japanese longline catch per unit effort (CPUE) may be a biased abundance estimator for albacore, bigeye and yellowfin tuna (*Thunnus alalunga*, *T. obesus* and *T. albacares*) in the tropical Pacific, one of the 13 regions we considered. The appropriate use of CPUE data is an important technical issue with substantial policy implications. We have therefore made every effort to check and correct the data for potential bias and to validate them against independent survey data. The results of our analyses<sup>3–7</sup> indicate that our main conclusions are still justified. We note that Hampton *et al.* also use uncorrected CPUE data, assuming that CPUE is proportional to abundance (see ref. 8, for example). As yet, they have not demonstrated a mechanism that could explain why the assumption of proportionality should break down.

We agree that changes in targeting, particularly the increase in the depth of hooks, have altered catchability; however, when the effects of depth are estimated<sup>3</sup>, the combined CPUE shows declines greater than those we estimated originally<sup>2</sup>. Other recent analyses of gear changes suggest that newer gear is twice as effective as older gear<sup>9</sup>, potentially obscuring continuing declines in stock abundance from CPUE data.

Hampton *et al.* discuss two cases. First,

they note that Taiwanese longline CPUE for albacore does not match the Japanese data (decline of about 50% compared with about 90%). However, the Taiwanese data may be misleading as they commenced a decade after the Japanese data and so fail to capture the start of industrialized fishing. Intense albacore fisheries were well developed in this region before the Taiwanese data were collected and had large effects on this species<sup>10</sup>. Thus, less of the decline is seen in the later data (the ‘shifting baseline’ syndrome<sup>11</sup>). The same issue applies to the tropical Pacific, where populations were exploited previously — those areas were therefore excluded from our analysis.

Second, the authors observe that for Pacific yellowfin and bigeye the initial decline in longline CPUE occurred under moderate fishing effort, whereas CPUE remained low and stable under later regimes of high fishing effort. We explained this pattern by an increase in fish productivity, caused by the decline of large predators. Both ecosystem models<sup>12,13</sup> and survey data<sup>5</sup> support this mechanism.

We note further that CPUE is measured in units of numbers of fish large enough to be harvested by longline hooks, which are very size-selective. Initially, there was an accumulated biomass of large fish. Thus, CPUE was very high, but dropped rapidly as those susceptible individuals were removed. CPUE declined as the fishery became completely dependent upon new fish recruiting to the population. Thus, related shifts in size distribution, CPUE and fishing effort may lead to rapid depletion of large individuals, even at relatively low initial fishing effort.

This simple mechanism is consistent with

the large (2–4-fold) reduction in the average size of large predatory fish, as seen from research survey data<sup>5</sup>. The same surveys revealed an 89% decline of large pelagic biomass in the tropical Pacific<sup>5</sup> that precisely matched our global estimate<sup>2</sup>. Hampton *et al.* imply that environmental effects are partly to blame for these changes. Although environmental factors can drive year-to-year variation<sup>8,14</sup>, they cannot explain long-term, worldwide declines, particularly as these coincide with the onset of industrialized fishing<sup>5,15</sup>.

We welcome the incentive from Hampton *et al.* to refine and critically evaluate abundance estimates derived from CPUE data. However, in following their call, either by correcting for potential biases or by analysing independent data sets, we find that our estimates of decline remain conservative.

**Ransom A. Myers, Boris Worm**

Biology Department, Dalhousie University, Halifax, Nova Scotia B3H 4J1, Canada  
e-mail: Ransom.Myers@dal.ca

doi:10.1038/nature03582

1. Hampton, J., Sibert, J. R., Kleiber, P., Maunder, M. N. & Harley, S. J. *Nature* **434**, doi:10.1038/nature03581 (2005).
2. Myers, R. A. & Worm, B. *Nature* **423**, 280–283 (2003).
3. Ward, P. & Myers, R. A. *Can. J. Fish. Aquat. Sci.* (in the press).
4. Ward, P., Myers, R. A. & Blanchard, W. *Fish. Bull.* **102**, 179–195 (2004).
5. Ward, P. & Myers, R. A. *Ecology* **86**, 835–847 (2005).
6. Ward, P. & Myers, R. A. *Fish. Oceanogr.* (in the press).
7. Myers, R. A. & Worm, B. *Phil. Trans. R. Soc. Lond. B* doi:10.1098/rstb.2004.1573 (2005).
8. Lehodey, P., Bertignac, M., Hampton, J., Lewis, A. & Picaut, J. *Nature* **389**, 715–718 (1997).
9. Stone, H. H. & Dixon, K. L. *Fish. Bull.* **99**, 210–216 (2001).
10. Otsu, T. & Sumida, R. F. *Fish. Bull.* **67**, 47–67 (1968).
11. Pauly, D. *Trends Ecol. Evol.* **10**, 430 (1995).
12. Cox, S. P. *et al.* *Can. J. Fish. Aquat. Sci.* **59**, 1736–1747 (2002).
13. Cox, S. P. *et al.* *Can. J. Fish. Aquat. Sci.* **59**, 1724–1735 (2002).
14. Myers, R. A. *Rev. Fish. Biol. Fish.* **8**, 285–305 (1998).
15. Worm, B. & Myers, R. A. *Nature* **429**, 15 (2004).

# Reduced sleep in *Drosophila Shaker* mutants

Chiara Cirelli<sup>1</sup>, Daniel Bushey<sup>1</sup>, Sean Hill<sup>1</sup>, Reto Huber<sup>1</sup>, Robert Kreber<sup>2</sup>, Barry Ganetzky<sup>2</sup> & Giulio Tononi<sup>1</sup>

<sup>1</sup>Department of Psychiatry, 6001 Research Park Blvd, University of Wisconsin Madison, Madison, Wisconsin 53719, USA

<sup>2</sup>Laboratory of Genetics, 445 Henry Mall, University of Wisconsin Madison, Madison, Wisconsin 53706, USA

**Most of us sleep 7–8 h per night, and if we are deprived of sleep our performance suffers greatly; however, a few do well with just 3–4 h of sleep—a trait that seems to run in families. Determining which genes underlie this phenotype could shed light on the mechanisms and functions of sleep. To do so, we performed mutagenesis in *Drosophila melanogaster*, because flies also sleep for many hours and, when sleep deprived, show sleep rebound and performance impairments. By screening 9,000 mutant lines, we found *minisleep* (*mns*), a line that sleeps for one-third of the wild-type amount. We show that *mns* flies perform normally in a number of tasks, have preserved sleep homeostasis, but are not impaired by sleep deprivation. We then show that *mns* flies carry a point mutation in a conserved domain of the *Shaker* gene. Moreover, after crossing out genetic modifiers accumulated over many generations, other *Shaker* alleles also become short sleepers and fail to complement the *mns* phenotype. Finally, we show that short-sleeping *Shaker* flies have a reduced lifespan. *Shaker*, which encodes a voltage-dependent potassium channel controlling membrane repolarization and transmitter release, may thus regulate sleep need or efficiency.**

Sleep occupies one-third of our life and an even larger proportion in infants<sup>1</sup>. It is present in virtually every animal species where it has been studied<sup>2</sup>, despite the danger of partial disconnection from the environment. Total sleep deprivation can be fatal<sup>3,4</sup>, and even partial deprivation of sleep has serious consequences on cognition, mood and health<sup>1,5</sup>. Sleep is tightly regulated in a homeostatic manner<sup>6</sup>: the longer one stays awake, the stronger the pressure to sleep, suggesting that sleep serves an essential function.

In most animals, the timing of sleep is gated by the circadian system, which ensures that sleep occurs at the appropriate time of day<sup>7</sup>. The circadian system is a prime example of how a complex behaviour can be dissected successfully by genetic means<sup>8,9</sup>. Starting with the discovery of the *period* gene in *Drosophila*<sup>10–12</sup>, mutagenesis studies in flies and mice have identified at least ten genes whose mutations critically affect circadian rhythms, and have made it possible to unravel the molecular mechanisms of the circadian clock<sup>8,9</sup>.

Animals continue to sleep many hours per day even if the timing of sleep is disrupted by anatomical or genetic lesions of the circadian clock<sup>13–16</sup>. Moreover, in arrhythmic animals sleep continues to be regulated homeostatically<sup>13–17</sup>, suggesting that an adequate amount of daily sleep is required to fulfil its biological functions. Although the key mechanisms controlling the circadian timing of sleep are well understood, those determining sleep amount remain unclear. Twin studies in humans<sup>18–20</sup> as well as animal studies<sup>21</sup> indicate that genetic factors can influence daily sleep duration. But how significant are such genetic factors, and can sleep need be affected strongly by a single gene? A few human cases of extremely short sleep have been reported, and some may show familial occurrence<sup>22–26</sup>. These reports suggest that it may be possible to identify genes that exert a major influence on the amount of sleep.

## Mutagenesis screening for short-sleeping flies

To identify single-gene mutations that affect daily sleep amount, we performed a mutagenesis screening in *Drosophila melanogaster*<sup>27</sup>. Sleep in fruitflies shares many similarities with mammalian sleep<sup>28–30</sup>. Flies show long periods of quiescence during which arousal thresholds are increased. Just as in mammals<sup>31</sup>, fly sleep is accompanied by changes in neural activity<sup>32</sup> and gene expression in the brain<sup>29</sup>. Moreover, sleep is abundant in young flies, decreases in older animals, and is modulated by stimulants and hypnotics<sup>29</sup>.

Finally, when flies are kept awake forcibly, their performance is impaired, and they show a compensatory increase in sleep duration and intensity the next day<sup>30</sup>, just as mammals do<sup>2</sup>.

To identify mutations that affect daily sleep amount in *Drosophila* we screened through approximately 6,000 ethylmethane sulphonate (EMS)-mutagenized *Drosophila* lines (X chromosome) and through a collection of about 3,000 lines carrying randomly inserted P and EP transposable elements<sup>33,34</sup>. We identified 15 lines having a mean daily amount of sleep at least 2 standard deviations less than the average for both males and females (Fig. 1a). EMS line E1174, one of the most extreme, was termed *minisleep* (*mns*) and designated for further study.

## Behavioural characterization of *minisleep* mutants

The *mns* phenotype maps to the X chromosome and is recessive. As shown in Fig. 1b, whereas wild-type flies and heterozygous female *mns* flies sleep for 9–15 h every day, both male and homozygous female *mns* flies slept for only 4–5 h. Sleep is defined here, as in previous work, as behavioural immobility lasting 5 min or more, which is consistently associated with increased arousal thresholds during both light and dark periods<sup>29,30</sup>. The short-sleeping phenotype of *mns* flies is not dependent on the specific duration criterion: after either 1 min or 5 min of immobility, both wild-type and *mns* flies showed a significant decrease in their ability to react to a stimulus (Fig. 1c). The decrease in daily sleep amount in *mns* flies was mainly due to a decrease in the duration of sleep episodes rather than in their number (Fig. 1d). Similarly, waking time in *mns* flies increased largely because waking episodes were longer rather than more frequent (Fig. 1d). The short-sleeping phenotype persisted under constant darkness, when sleep amounts were even lower than under light/dark conditions (Fig. 1e). Moreover, under constant darkness *mns* flies maintained a rhythmic modulation of locomotor activity with a period of ~24 h (Fig. 1f; see also Supplementary Fig. 1 for activity profiles of individual flies). Thus, the *mns* mutation affects the amount but not the circadian regulation of sleep.

We next examined the response of *mns* flies to sleep deprivation. When sleep deprived, wild-type flies show an increase in sleep duration and intensity, just as mammals do<sup>30</sup>, suggesting that sleep serves some function that is homeostatically regulated<sup>2</sup>. Both male and female *mns* flies also showed a significant increase in sleep duration in response to 24 h of sleep deprivation, and they recovered

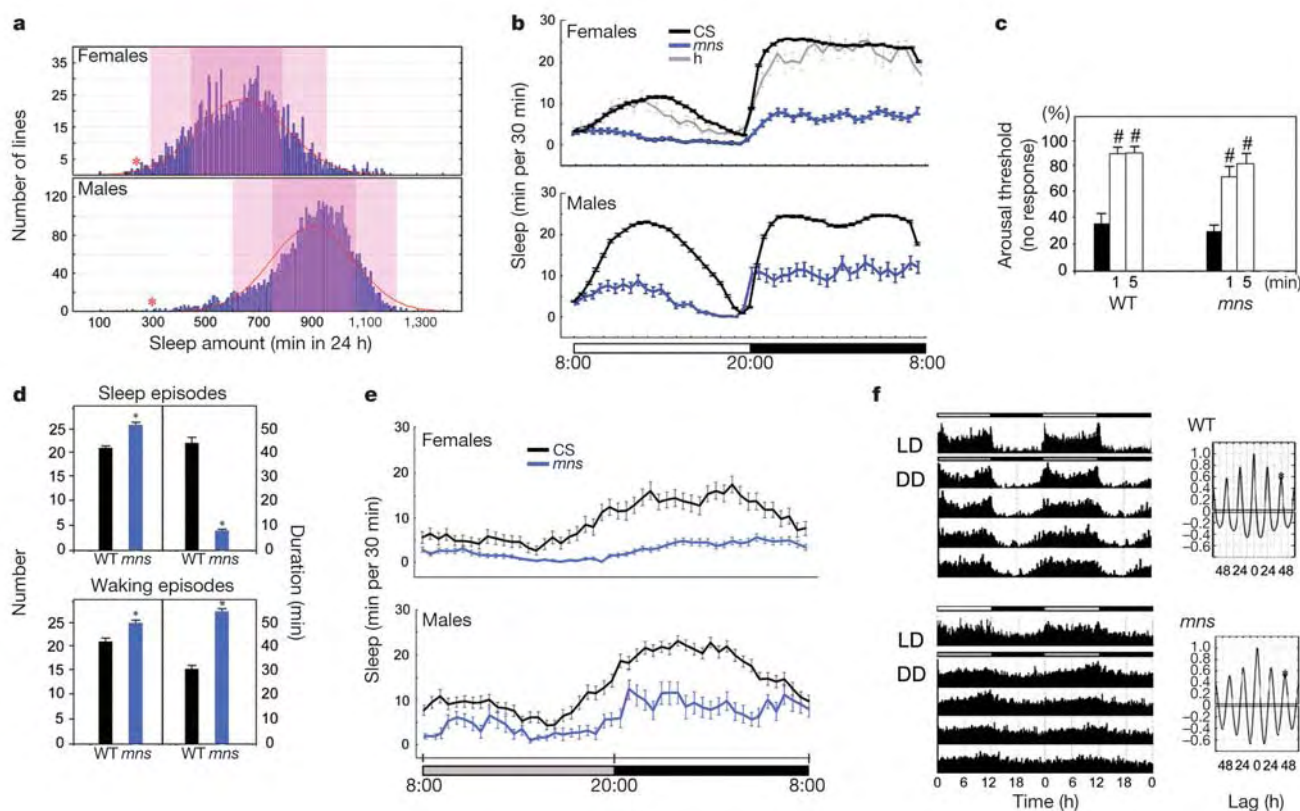
a similar percentage of the sleep they lost (Fig. 2a). Moreover, their sleep became more intense, as indicated by a decrease in the number of brief awakenings, an increase in the duration of sleep episodes, and a trend for increase in arousal threshold (Fig. 2b). However, these changes were significantly smaller than in wild-type flies (Fig. 2b). Thus, sleep in *mns* flies is homeostatically regulated, but intensity changes seem to be less prominent than in wild-type flies.

Adult *mns* flies are homozygous viable and fertile with no obvious behavioural or developmental phenotype. When awake, *mns* flies moved and explored normally, indicating that the mutation affects neither their general locomotor ability nor makes them hyperactive (Fig. 2c). *mns* flies also performed normally in several behavioural tasks, including the geotactic response and the ability to climb at high temperature (not shown). We also examined *mns* flies for their ability to respond to complex and thermal stimuli<sup>30</sup>. As shown in Fig. 2d, the escape response of awake *mns* flies during baseline was

within the range observed in wild-type flies. Notably, however, whereas the performance of wild-type flies is markedly reduced after sleep deprivation<sup>30</sup>, *mns* flies were not impaired (Fig. 2d). Altogether, *mns* flies have normal activity levels and normal waking performance in several tasks, their sleep is of short duration rather than fragmented, their homeostatic response to forced wakefulness is present but reduced in its intensity, and they are resistant to damaging effects of sleep deprivation on performance. Thus, *mns* flies may have a reduced sleep need or may be able to sleep more efficiently, although we cannot rule out that they may have some difficulty in obtaining enough sleep. Moreover, we cannot exclude that *mns* flies may be defective in tasks that were not examined here.

### Molecular characterization of the *minisleep* mutation

What is responsible for the *mns* phenotype? EMS mutagenesis can cause point mutations in any region of the genome with little bias,



**Figure 1** Sleep in *mns* flies. **a**, Distribution of daily sleep amounts in about 9,000 mutant lines (16 flies per line,  $\geq 3$  experiments per line). Shaded areas show one and two standard deviations from the mean (mean  $\pm$  s.d.: females,  $624 \pm 167$ ; males,  $910 \pm 155$ ; daily sleep amount is  $>2$  s.d. lower than the mean in females of 36 lines and in males of 181 lines). In almost all lines female flies sleep less than male flies, in agreement with previous studies<sup>30,42</sup>. Red asterisks indicate *mns* flies. **b**, Daily time course (30-min intervals) of the amount of sleep in wild-type Canton-S (CS) and *mns* flies. Curves connect mean values  $\pm$  s.e.m. (min of sleep per 24 h: CS females ( $n = 63$ ) =  $664 \pm 17$ ; CS males ( $n = 55$ ) =  $923 \pm 21$ ; *mns* heterozygous (h) females ( $n = 50$ ) =  $564 \pm 32$ ; *mns* homozygous females ( $n = 60$ ) =  $247 \pm 22$ ; *mns* males ( $n = 58$ ) =  $297 \pm 34$ ). The white and black bars under the x axis indicate light and dark periods, respectively. **c**, Arousal threshold differences between periods of activity and immobility (mean  $\pm$  s.e.m. for the entire 12-h dark period). Arousal threshold is measured as the percentage of flies that did not show an escape response after the delivery of a complex stimulus of low intensity<sup>30</sup>. Only a minority of wild-type (WT) and *mns* flies do not respond if they had been active during the minute before the stimulus was delivered (filled columns); the number of non-responsive flies increases significantly when flies are stimulated after a period of immobility of at least 5 min (15 female plus 15 male flies per line; hash symbol indicates  $P < 0.01$ , paired *t*-test). During the dark period the

arousal threshold is also significantly increased after 1 min of immobility. **d**, Duration (in min) and number of sleep and waking episodes during 24 h of baseline recording in wild-type and *mns* flies (mean  $\pm$  s.e.m., 16 female plus 16 male flies per line; asterisk indicates  $P < 0.05$ , *t*-test). **e**, Daily sleep amount (mean  $\pm$  s.e.m., in min) after 7 days of constant darkness (DD): CS females =  $450 \pm 44$ ; CS males =  $651 \pm 21$ ; *mns* females =  $153 \pm 24$ ; *mns* males =  $268 \pm 37$  ( $n > 50$  flies per line per gender). In DD conditions *mns* flies of both genders still sleep less than wild-type flies ( $P < 0.05$ , Student's *t*-test). All flies tend to sleep less under DD conditions than under light/dark (LD) conditions (compare **b** with **e**), but the difference is significant only for CS flies ( $P < 0.05$ , paired *t*-test). The grey bar under the plots represents subjective day, the black bar represents subjective night. **f**, The left panels show average activity histograms in wild-type (top panel) and *mns* (bottom panel) female flies over 2 days in LD and 5 days in constant darkness (DD). Actograms in DD are double plotted ( $>100$  female flies per line). The right panels show an autocorrelation analysis of locomotor behaviour as calculated according to ref. 49 in wild-type and *mns* flies kept in DD for 7 consecutive days. The asterisks indicate the rhythmicity index (RI), a measure of the strength of the activity rhythm (RI: wild type = 0.54, *mns* = 0.52). The estimated period is 24.0 h in wild-type flies and 24.1 h in *mns* flies. Autocorrelation analysis for individual *mns* flies indicates that  $>90\%$  are rhythmic, similar to wild-type flies (data not shown).



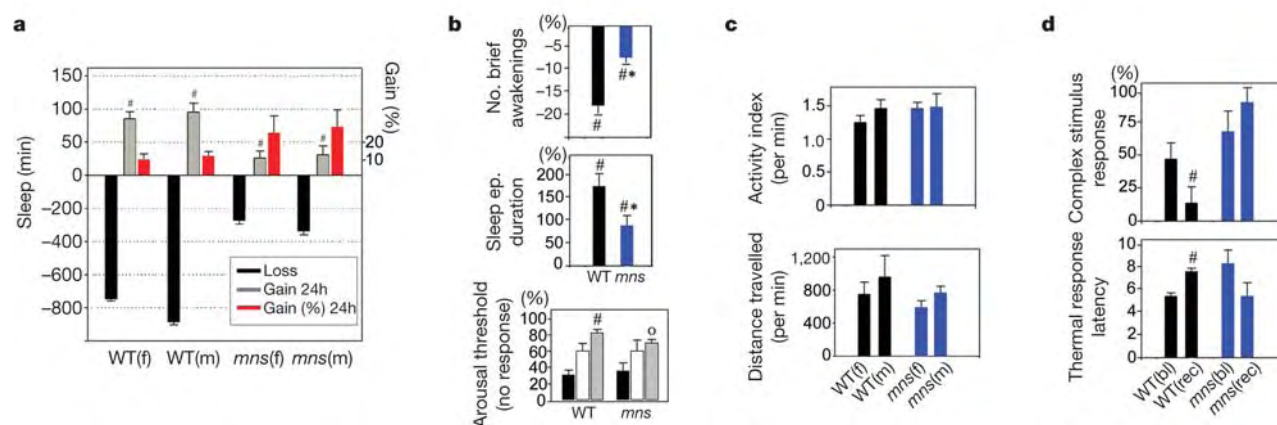
but it may be difficult to identify where the mutation has occurred. While performing the behavioural characterization of *mns* flies, we noticed that they exhibited a transient shaking of the legs and scissoring of the wings when recovering from diethyl ether anaesthesia. Four genes on the X chromosome, when mutated, can result in a shaking phenotype: *Shaker*, *ether a go-go*, *Hyperkinetic* and *shaking B*. The shaking phenotype of *mns* flies was dominant, persisted in amputated legs, and was not seen when intact flies were exposed to chloroform, halothane, isoflurane or enflurane, strongly suggesting a mutation of the *Shaker* locus. Genetic analysis confirmed that the shaking phenotype maps to the *Shaker* locus. Crucially, the short-sleeping phenotype mapped to the same locus, and no recombinants were identified that were short sleepers but not shakers (Supplementary Fig. 2a, b). Thus, a mutation in the *Shaker* (*Sh*) locus appears to cause both phenotypes. In this case, it should be possible to rescue the *mns* phenotype, which is recessive, by using a *Sh*<sup>+</sup> transgene or duplication. Because the *Sh* locus is very large and produces many alternative transcripts<sup>35</sup>, we tested a duplication that included the *Sh*<sup>+</sup> locus (*Dp(1;3)JC153*). As shown in Supplementary Fig. 2c, when one or more copies of *Sh*<sup>+</sup> were present, *mns* flies reverted as expected to a normal sleep phenotype.

The *mns* phenotype was isolated after mutating male flies from a wild-type CS stock. To determine whether the *mns* phenotype depends on genetic background, heterozygous (*mns/Sh*<sup>+</sup>) shaking females were backcrossed consecutively, over several generations, to

males of a different control stock, w<sup>1118</sup>. The shaking and non-shaking male progeny were isolated from each generation and tested for the daily amount of sleep. Over five generations, shaking flies had a short-sleeping phenotype ( $363 \pm 35$ ,  $n = 44$ ) whereas non-shaking flies had a wild-type phenotype ( $858 \pm 31$ ,  $n = 58$ ). Thus, *mns* produces a robust short-sleeping phenotype regardless of genetic background.

The *Shaker* locus encodes the  $\alpha$ -subunit of a tetrameric voltage-dependent potassium channel that controls membrane repolarization after action potentials and presynaptic transmitter release<sup>36</sup>. Homologous channels in vertebrates have similar properties<sup>37</sup>. The shaker phenotype is due to a peripheral neuromuscular effect, but central effects of *Shaker* mutations are well documented, including a reduced sensitivity to anaesthetics such as halothane<sup>38</sup>. *Shaker* isoforms are expressed mostly in the brain, especially in integrative regions such as the mushroom bodies, and most strongly in the neuropil in association with axons and synaptic terminals<sup>39</sup>. All isoforms share a core region that includes the S1 domain. Domains S1–S4 constitute the voltage-sensing module, whereas domains S5–S6 form the selectivity pore (Fig. 3a, b).

We sequenced transcripts and genomic DNA of the entire *Shaker* coding region in the *mns* line and in two normal-sleeping EMS lines generated at the same time as E1174. The only sequence difference that caused an amino acid substitution in *mns* flies was a point mutation (C to T) in exon 9, resulting in a threonine to isoleucine substitution (Fig. 3b). This substitution of a polar amino acid with



**Figure 2** Response to sleep deprivation and measures of performance in *mns* flies. **a**, Increase in sleep duration after sleep deprivation ( $n > 70$  flies per line per gender; five experiments). Black columns, sleep lost (in min) during 24 h of sleep deprivation (sleep lost was  $>90\%$  of baseline sleep in all experiments); grey columns, sleep gain (the number of minutes flies overslept relative to baseline during the first 24 h after sleep deprivation; hash indicates  $P < 0.05$ , paired  $t$ -test). The amount of sleep recovered, expressed as percentage of sleep lost (red columns, right y axis) is not significantly different in wild-type (females and males) and *mns* flies (mean  $\pm$  s.e.m., wild-type females (WT(f)) =  $11 \pm 3\%$ ; WT(m) =  $13 \pm 3\%$ ; *mns*(f) =  $26 \pm 13\%$ ; *mns*(m) =  $29 \pm 12\%$ ;  $P = 0.29$ ,  $t$ -test). None of these flies was heat-shocked before sleep deprivation. **b**, Increase in sleep intensity after sleep deprivation. In all flies the number of brief awakenings (top panel) is significantly reduced during the light period after sleep deprivation relative to baseline, whereas the duration of the sleep episodes (middle panel) is significantly increased (hash indicates  $P < 0.05$ , paired  $t$ -test,  $n = 32$  flies per line). In both cases the change is significantly smaller in *mns* flies relative to wild-type flies (asterisk,  $P < 0.05$ ,  $t$ -test). Arousal threshold was measured as in Fig. 1c (bottom panel). Black columns represent the percentage of no response in flies that had been moving the minute before the stimulus was delivered; white and grey columns refer to flies that had been immobile for 5 min (sleeping flies). Arousal thresholds (the percentage of non-responsive flies) are higher during recovery sleep after sleep deprivation (grey columns) relative to baseline sleep (white columns) in both wild-type and *mns* flies, but the change is significant only in wild-type flies ( $n = 32$  male flies per line;

hash indicates  $P < 0.05$ , paired  $t$ -test) and not in *mns* flies ( $0$ ,  $P = 0.08$ ). Values are mean  $\pm$  s.e.m. for the first 6 h of the light period after sleep deprivation. **c**, The top panel shows locomotor activity measured in the DAMS monitor as activity index (that is, the number of beam crossings per min;  $n = 16$  female and 16 male flies per line, light period). Values include only awake flies (mean  $\pm$  s.e.m.). The bottom panel shows locomotor behaviour measured in the heat box during the last 10 min of the adaptation period preceding the delivery of the thermal stimulus (all flies were awake during that period). The distance travelled per min is measured in arbitrary units (mean  $\pm$  s.e.m.) ( $n = 25$  flies per line). **d**, Assessment of performance before and after sleep deprivation. Top panel: the response to a complex stimulus is measured as the percentage increase in the number of beam crossings during the minute after the delivery of the stimulus relative to the minute before the stimulation. In wild-type flies, but not in *mns* flies, the increase is significantly reduced during recovery (rec) after sleep deprivation relative to baseline (bl). All flies had been active (awake) during the minute before the delivery of the stimulus. Values are mean  $\pm$  s.e.m. averaged for the entire light period (one stimulus per hour;  $n = 32$  flies per line, hash indicates  $P < 0.05$ , paired  $t$ -test). Bottom panel: the response to a thermal stimulus is measured as the latency (in seconds) to beam crossing after heat was applied to the side of the chamber housing the flies (mean  $\pm$  s.e.m.,  $n = 32$  flies per line). The escape response after sleep deprivation worsens (latency increases) in wild-type flies but not in *mns* flies. Each fly was tested once during the first 2 h after the end of sleep deprivation and at the corresponding time of day during baseline (hash indicates  $P < 0.05$ , paired  $t$ -test).

a highly hydrophobic one occurs at the extracellular end of S1 (Fig. 3c). The mutated threonine residue is extremely well conserved from *Aplysia* to human (Fig. 3c), suggesting that it has some important role. Moreover, it appears to be in close proximity to the voltage-sensing S4 domain as well as to the pore domain<sup>40</sup>, and certain amino acid substitutions in the S1–S2 loop are known to produce changes in its voltage dependency<sup>41</sup>. Northern blot analysis found no difference in the amount and distribution of *Shaker* transcripts between wild-type and *mns* flies, and confirmed that most *Shaker* messenger RNA is expressed in the fly head (Fig. 3d).

### Sleep and other *Shaker* alleles

Is the short-sleeping phenotype specific for the *mns* mutation, or does it occur in other *Shaker* mutants and in other ion channel mutants? To find out, we measured sleep amount for other *Shaker* alleles as well as for several alleles of voltage-gated K<sup>+</sup> channels (*ether a go-go*, *seizure*), voltage-gated Na<sup>+</sup> channels (*paralytic*, *tipE*) and calcium-activated K<sup>+</sup> channels (*slowpoke*). We found that all non-*Shaker* ion channel mutants tested were normal sleepers (not shown), whereas *Shaker* flies of line *Sh*<sup>102</sup>, a characterized null allele, slept at least 2 standard deviations less than average for both genders (Fig. 4a).

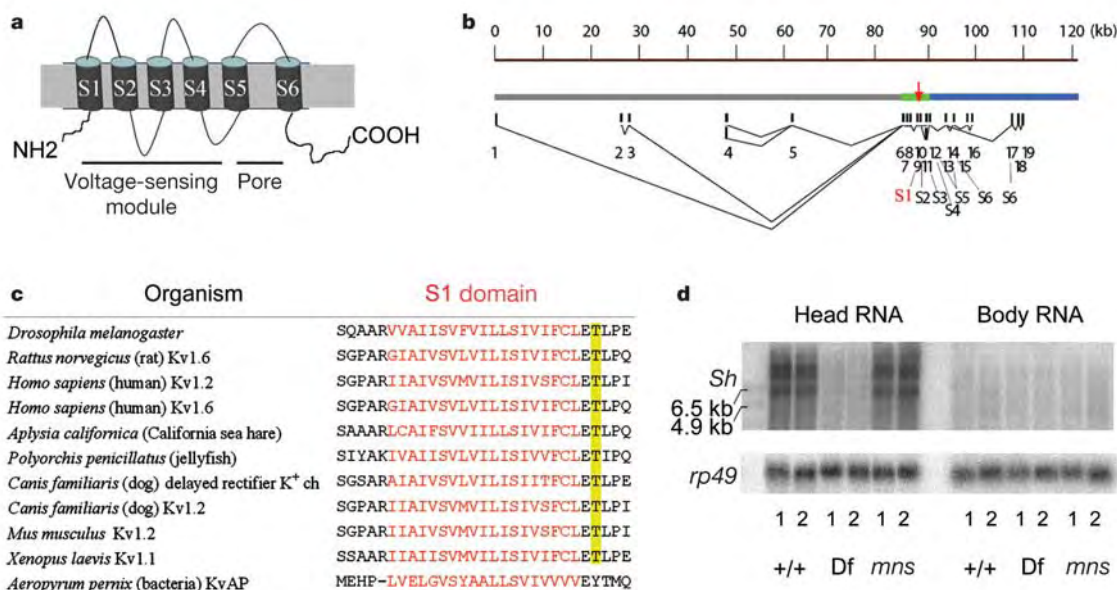
Although these findings suggest that the *Shaker* gene may have a special role in controlling *Drosophila* sleep, other *Shaker* lines, including the null alleles *Sh*<sup>133</sup> and *Sh*<sup>M</sup>, had normal amounts of sleep. Why is this the case? Even in *mns* flies, the short-sleeping phenotype shows incomplete penetrance, suggesting that its expression may be sensitive to modifiers. We reasoned that if modifiers were suppressing the short-sleeping phenotype then they could be removed by consecutively crossing the mutated *Sh* alleles with wild-type (*Sh*<sup>+</sup>) lines and selecting for the shaking phenotype. Just as hypothesized, null alleles *Sh*<sup>102</sup>, *Sh*<sup>M</sup> and *Sh*<sup>133</sup> became extreme short sleepers after outcrossing, irrespective of the stock background (Fig. 4a). Moreover, irrespective of stock background, all these null alleles failed to complement the short-sleeping phenotype of *mns* flies (Supplementary Fig. 2d), proving that *mns* is

a bona fide *Shaker* allele. Other *Shaker* alleles, including the characterized hypomorph *Sh*<sup>5</sup>, also became short sleepers after outcrossing (Fig. 4a). Thus, in the absence of modifiers, a dysfunctional *Shaker* gene appears to lead to a short-sleeping phenotype regardless of the particular mutation.

Conceivably, modifiers may have accumulated in long-standing stocks through selective pressure. Indeed, when we calculated longevity profiles in *mns* flies, we found a reduction in lifespan as compared with the control strain (Fig. 4b). We then examined outcrossed *Shaker* alleles (*Sh*<sup>102</sup> or *Sh*<sup>M</sup>) and found that short-sleeping, shaking flies also had reduced lifespan compared with non-short-sleeping, non-shaking siblings (Fig. 4b). Non-outcrossed flies carrying the same alleles, which slept more but were still shakers, had a normal lifespan (Fig. 4b). These results suggest that the short-sleeping phenotype as such may affect longevity, although outcrossed flies that became short sleepers also tended to shake more strongly. Intriguingly, another non-shaking, short-sleeping line identified through our large-scale mutagenesis has normal performance but reduced lifespan (data not shown), and reducing sleep amount in wild-type and *cycle*<sup>01</sup> flies through dark-rearing also reduces lifespan<sup>42</sup>.

### Implications for mammalian sleep

A growing body of evidence in humans and animals indicates that several aspects of sleep may be, in part, under genetic control<sup>21,43</sup>. Here we have shown that a single gene mutation can produce an extreme short-sleeping phenotype, and we have mapped the phenotype to the *Shaker* locus. There are previous reports of slight alterations of sleep amount in mouse mutants<sup>21</sup>, but none exceeds the range of variation observed among wild-type strains. We<sup>27</sup> and others<sup>44</sup> have also identified a few *Drosophila* lines with reduced sleep, although none as extreme as for both genders of the *mns* line. It is possible that the *mns* mutation, by affecting an ion channel that controls membrane repolarization, may be close to the core cellular mechanisms of sleep. In mammals, potassium channels are involved in the generation of sleep rhythms<sup>45–47</sup>. It is not known whether

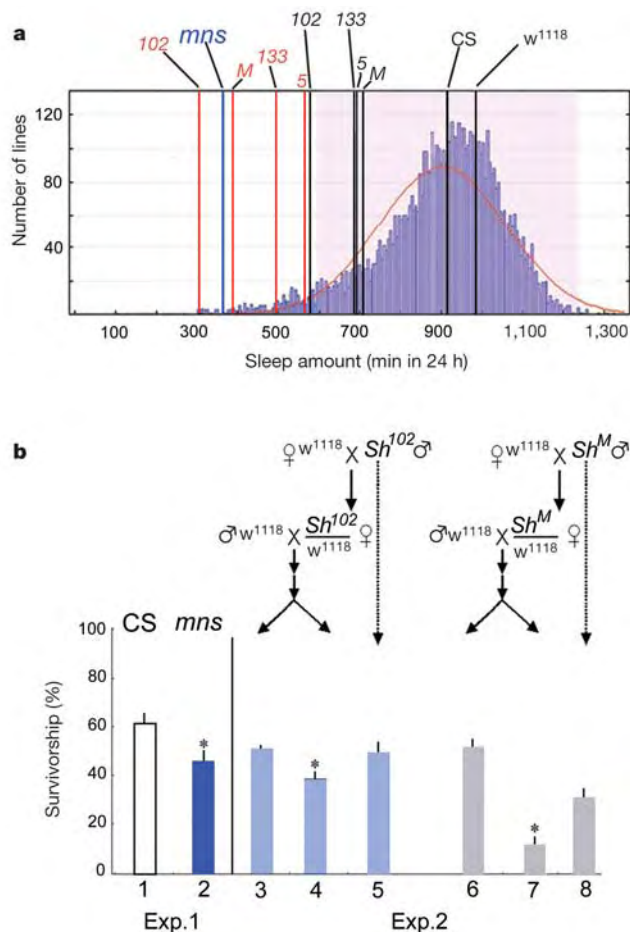


**Figure 3** The *Shaker* channel and the *mns* mutation. **a**, The  $\alpha$ -subunit of the *Shaker* channel includes six transmembrane segments: S1–S4 form the voltage-sensor module, S5–S6 form the pore region. **b**, Schematic representation of the *Shaker* transcription unit with 19 exons. The grey bar indicates the amino-terminal variable region; the green bar indicates the common central region; and the blue bar indicates the carboxy-terminal variable region. The red arrow indicates the approximate location of the *mns* mutation.

**c**, Sequence alignment of the S1 domain. The threonine residue is conserved between *Shaker* homologues in different species. **d**, *Shaker* transcripts from fly heads and bodies. The probe was a fragment of 550 base pairs spanning exons 9 and 10. To check that equal amounts of RNA were being compared, blots were re-probed with probes for *rp49*. Df, *Shaker* deficiency line B55D/W32P.

human extreme short sleepers have mutations in depolarization or voltage-dependent potassium channels. However, in Morvan's syndrome, a rare autoimmune disorder with central nervous system symptoms, marked sleeplessness has been associated with autoantibodies against voltage-dependent potassium channels that may have crossed the blood–brain barrier<sup>48</sup>. The finding that a point

mutation in a voltage-dependent potassium channel produces an extreme short-sleeping phenotype with preserved performance is relevant for at least three reasons: it shows that sleep, a complex phenotype, can be affected powerfully by a single gene; it opens the way to identifying specific mechanisms by which sleep performs its functions at the cellular level; and it suggests that it may be possible to develop molecular tools that prolong wakefulness and promote short periods of highly restorative sleep. □



**Figure 4** Effects of outcrossing on daily sleep amount and longevity. **a**, Distribution of daily sleep amounts in male flies of approximately 9,000 mutant lines and in several *Shaker* alleles. Black and red lines refer to sleep amounts before and after outcrossing to *w<sup>1118</sup>*, respectively. Sleep amount in *w<sup>1118</sup>* and CS flies is shown for reference. The blue line indicates *mns* flies outcrossed to *w<sup>1118</sup>*. For each line at least 30 flies were tested in at least two independent experiments. The severity of the short-sleeping phenotype among different alleles, with *Sh<sup>102</sup>* being the strongest and *Sh<sup>5</sup>* the least extreme, is consistent with the severity of the changes in the rapidly inactivating A-type  $K^+$  current mediated by *Shaker*<sup>50</sup>. Outcrossing to CS produced results similar to those shown for *w<sup>1118</sup>*. Consistent with the fact that CS flies sleep less than *w<sup>1118</sup>* flies, and that *mns* flies are more extreme short sleepers on a CS background than on a *w<sup>1118</sup>* background, outcrossing to CS produced stronger short sleepers than outcrossing to *w<sup>1118</sup>* (min per day, mean  $\pm$  s.e.m.: *Sh<sup>102</sup>* =  $121 \pm 23$ ; *Sh<sup>133</sup>* =  $268 \pm 33$ ; *Sh<sup>5M</sup>* =  $234 \pm 52$ ; *Sh<sup>5</sup>* =  $471 \pm 34$ ). Shaded area is 2 s.d. from the mean. **b**, Percentage of surviving flies at 21 °C at eight weeks of age. Experiment 1 compares male CS flies (group 1) to male *mns* flies in a CS background (group 2; >350 flies per line, three independent experiments per line; mean  $\pm$  s.e.m.). The crossing scheme for experiment 2 is depicted above the chart and the arrows indicate the progeny tested (150 flies per line). After three generations of outcrossing, male flies inheriting the *Sh<sup>102</sup>* (group 4) or the *Sh<sup>5M</sup>* (group 7) allele were short sleepers and shakers and had a reduced lifespan compared with their non-short-sleeping and non-shaking male siblings (*Sh<sup>+</sup>*, groups 3 and 6). Lifespan was also reduced in groups 4 and 7 compared with the original non-outcrossed stocks *Sh<sup>102</sup>* (group 5) and *Sh<sup>5M</sup>* (group 8), which slept more but were still shakers. Asterisk indicates a significant decrease in survivorship in group 2 versus group 1, in group 4 versus groups 3 and 5, and in group 7 versus groups 6 and 8 ( $P < 0.05$ , Student's *t*-test).

## Methods

### Animals

Flies (1–2 weeks old) were cultured and tested at 21 °C, 68% humidity, on yeast, dark corn syrup and agar food. EMS mutagenesis screening for X-linked mutations was carried out on a Canton-S background.

### Locomotor activity, sleep and measures of sleep intensity

Experiments included one day of adaptation, two baseline days, one sleep deprivation day and two recovery days after sleep deprivation. At the beginning of the experiment, individual flies were placed in the *Drosophila* Activity Monitor System (DAMS, Trikinetics) inside glass tubes with enough food for 1 week of recording. Monitors were housed inside environmental chambers (ThermoForma) where temperature and humidity were kept constant. Data analysis was performed by custom-designed software developed in our laboratory<sup>30</sup> and based on Statistica (StatSoft). The data were further analysed using Matlab (Mathworks). Sleep and wakefulness were determined for consecutive 1-min periods. Wakefulness was defined as any period of at least 1 min characterized by activity ( $\geq 1$  count per min). On the basis of previous work<sup>29,30</sup>, sleep was defined as any period of uninterrupted behavioural immobility (0 counts per min) lasting  $> 5$  min. The duration of sleep episodes was calculated by counting the number of consecutive 1-min periods of sleep. Brief awakenings were defined as 1-min periods with at least one count preceded and followed by 1-min periods with no counts.

### Escape response to a complex stimulus

Flies were exposed to a complex stimulus consisting of a combination of noise and vibration. Flies remained inside a DAMS monitor, which was inserted into a custom-made frame specifically designed for the test<sup>30</sup>. The stimulus was produced by a flap vigorously pushed for a few seconds against the glass tubes housing the flies. Such stimulus was delivered once every hour at either side of the tubes via a computer-controlled motor, and flies were tested for a total of 48 h, including one baseline day and the first recovery day after sleep deprivation. Previous studies have shown that most flies move away from the stimulus (and by doing so cross an infrared beam) if, before its delivery, they had been actively moving around. By contrast, most flies do not show an escape response if they had been immobile for 5 min before the stimulus was delivered<sup>30</sup>. Thus, the percentage of non-responsive flies is used as a measure of the arousal threshold to distinguish awake flies from sleeping flies.

### Escape response to heat

Single flies were placed inside a heat box where position and movement of the fly were recorded and displayed online<sup>30</sup>. Flies were first adapted to the chamber for 30 min. Temperature on either side of the chamber was then alternately increased by 4 °C every minute from a baseline value to 44 °C (1 min each at 24, 28, 32, 36, 40 and 44 °C). The latency to crossing the infrared beam (that is, the time a fly needed to move to the cooler side of the chamber) was measured for each temperature step. Latencies for all temperature steps were averaged for each fly. Most flies took  $< 8$  s to move to the cold side of the chamber. Thus, on average, the total heat exposure for each fly lasted for 30–40 s. Flies were tested during the first 2 h after the end of sleep deprivation and at the same time of day during baseline. Pilot studies showed that the response to heat does not habituate in flies tested during two consecutive baseline days.

### Sleep deprivation

During sleep deprivation, flies remained in the DAMS monitor, which was placed vertically inside a framed box able to rotate along its major axis under the control of a motor<sup>30</sup>. The box could rotate 180° clockwise or anticlockwise (2–3 r.p.m.). At the nadir of each rotation, the monitor was dropped 1 cm. This caused the flies to fall from their current position to the bottom of the tube. Previous studies<sup>30</sup> had shown that this method is effective in reducing total sleep time by  $> 90\%$ . Since locomotor activity during sleep deprivation was continuously recorded, the extent of sleep loss could be calculated for each individual fly.

### Lifespan

Lifespan was measured by collecting newly enclosed flies, placing them in single tubes in a DAMS monitor, and recording their locomotor behaviour at 21 °C until death was noted. Flies were transferred to new tubes with fresh food every week. Lifespans were generated by calculating the percentage of survivorship daily and plotting this as a function of time in days.

### Statistical analysis

Two-way analysis of variance (ANOVA) tests with factors 'day' (for example, baseline versus recovery) and 'line' were used to analyse the data. Contrasts were tested by post-hoc *t*-test if the main factor or interaction reached significance.



## RNA extraction and analysis

Total RNA was isolated by using Trizol (GIBCO-BRL) according to the manufacturer's instructions. Final RNA concentrations were determined spectrophotometrically. RNA was transferred onto nylon membranes. <sup>32</sup>P-labelled probes were prepared by random priming of a purified polymerase chain reaction-amplified fragment using Megaprime DNA labelling system (Amersham Biosciences). Prehybridization (1 h) and hybridization (overnight) were performed at 42 °C in Hybrisol (Serologicals Corporation). The membranes were washed three times at room temperature in 2 × SSC/1% SDS, two times 30 min each at 50 °C in 2 × SSC/1% SDS, and one time at 60 °C for 30 min in 0.5 × SSC/1% SDS before being exposed to a phosphorimager (Molecular Dynamics).

Received 30 January; accepted 23 February 2005; doi:10.1038/nature03486.

1. Horne, J. A. *Why we Sleep. The Functions of Sleep in Humans and Other Mammals* (Oxford Univ. Press, Oxford, 1988).
2. Tobler, I. in *Principles and Practice of Sleep Medicine* (eds Kryger, M. H., Roth, T. & Dement, W. C.) 72–81 (W. B. Saunders, Philadelphia, 2000).
3. Rechtschaffen, A., Gilliland, M. A., Bergmann, B. M. & Winter, J. B. Physiological correlates of prolonged sleep deprivation in rats. *Science* **221**, 182–184 (1983).
4. Shaw, P. J., Tononi, G., Greenspan, R. J. & Robinson, D. F. Stress response genes protect against lethal effects of sleep deprivation in *Drosophila*. *Nature* **417**, 287–291 (2002).
5. Van Dongen, H. P., Maislin, G., Mullington, J. M. & Dinges, D. F. The cumulative cost of additional wakefulness: dose-response effects on neurobehavioral functions and sleep physiology from chronic sleep restriction and total sleep deprivation. *Sleep* **26**, 117–126 (2003).
6. Borbely, A. A. & Achermann, P. Sleep homeostasis and models of sleep regulation. *J. Biol. Rhythms* **14**, 557–568 (1999).
7. Dijk, D. J. & Lockley, S. W. Integration of human sleep-wake regulation and circadian rhythmicity. *J. Appl. Physiol.* **92**, 852–862 (2002).
8. Stanewsky, R. Genetic analysis of the circadian system in *Drosophila melanogaster* and mammals. *J. Neurobiol.* **54**, 111–147 (2003).
9. Lowrey, P. L. & Takahashi, J. S. Mammalian circadian biology: elucidating genome-wide levels of temporal organization. *Annu. Rev. Genom. Hum. Genet.* **5**, 407–441 (2004).
10. Konopka, R. J. & Benzer, S. Clock mutants of *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **68**, 2112–2116 (1971).
11. Bargiello, T. A., Jackson, F. R. & Young, M. W. Restoration of circadian behavioural rhythms by gene transfer in *Drosophila*. *Nature* **312**, 752–754 (1984).
12. Zehring, W. A. *et al.* P-element transformation with period locus DNA restores rhythmicity to mutant, arrhythmic *Drosophila melanogaster*. *Cell* **39**, 369–376 (1984).
13. Naylor, E. *et al.* The circadian clock mutation alters sleep homeostasis in the mouse. *J. Neurosci.* **20**, 8138–8143 (2000).
14. Wisor, J. P. *et al.* A role for cryptochromes in sleep regulation. *BMC Neurosci.* **3**, 20 (doi:10.1186/1471-2202-3-20) (2002).
15. Kopp, C., Albrecht, U., Zheng, B. & Tobler, I. Homeostatic sleep regulation is preserved in mPer1 and mPer2 mutant mice. *Eur. J. Neurosci.* **16**, 1099–1106 (2002).
16. Shiromani, P. J. *et al.* Sleep rhythmicity and homeostasis in mice with targeted disruption of mPeriod genes. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **287**, R47–R57 (2004).
17. Easton, A., Meerlo, P., Bergmann, B. & Turek, F. W. The suprachiasmatic nucleus regulates sleep timing and amount in mice. *Sleep* **27**, 1307–1318 (2004).
18. Heath, A. C., Kendler, K. S., Eaves, L. J. & Martin, N. G. Evidence for genetic influences on sleep disturbance and sleep pattern in twins. *Sleep* **13**, 318–335 (1990).
19. Linkowski, P. EEG sleep patterns in twins. *J. Sleep Res.* **8** (suppl. 1), 11–13 (1999).
20. Partinen, M., Kaprio, J., Koskenvuo, M., Putkonen, P. & Langinvainio, H. Genetic and environmental determination of human sleep. *Sleep* **6**, 179–185 (1983).
21. Tafti, M. & Franken, P. Invited review: genetic dissection of sleep. *J. Appl. Physiol.* **92**, 1339–1347 (2002).
22. Meddis, R., Pearson, A. J. & Langford, G. An extreme case of healthy insomnia. *Electroencephalogr. Clin. Neurophysiol.* **35**, 213–214 (1973).
23. Jones, H. S. & Oswald, I. Two cases of healthy insomnia. *Electroencephalogr. Clin. Neurophysiol.* **24**, 378–380 (1968).
24. Webb, W. B. Individual differences in sleep length. *Int. Psychiatry Clin.* **7**, 44–47 (1970).
25. Stuss, D. & Broughton, R. Extreme short sleep: personality profiles and a case study of sleep requirement. *Waking Sleep* **2**, 101–105 (1978).

26. Schenck, C. H. & Mahowald, M. W. Severe, childhood-onset, idiopathic, life-long insomnia responding selectively to opiate therapy: case report with 19 year follow-up. *Sleep Med.* **2**, 531–536 (2001).
27. Cirelli, C. Searching for sleep mutants of *Drosophila melanogaster*. *Bioessays* **25**, 940–949 (2003).
28. Hendricks, J. C. *et al.* Rest in *Drosophila* is a sleep-like state. *Neuron* **25**, 129–138 (2000).
29. Shaw, P. J., Cirelli, C., Greenspan, R. J. & Tononi, G. Correlates of sleep and waking in *Drosophila melanogaster*. *Science* **287**, 1834–1837 (2000).
30. Huber, R. *et al.* Sleep homeostasis in *Drosophila melanogaster*. *Sleep* **27**, 628–639 (2004).
31. Cirelli, C., Gutierrez, C. M. & Tononi, G. Extensive and divergent effects of sleep and wakefulness on brain gene expression. *Neuron* **41**, 35–43 (2004).
32. Nitz, D. A., van Swinderen, B., Tononi, G. & Greenspan, R. J. Electrophysiological correlates of rest and activity in *Drosophila melanogaster*. *Curr. Biol.* **12**, 1934–1940 (2002).
33. Rorth, P. *et al.* Systematic gain-of-function genetics in *Drosophila*. *Development* **125**, 1049–1057 (1998).
34. Spradling, A. C. *et al.* The Berkeley *Drosophila* Genome Project gene disruption project: Single P-element insertions mutating 25% of vital *Drosophila* genes. *Genetics* **153**, 135–177 (1999).
35. Sutcliffe, J. G. & Milner, R. J. Alternative mRNA splicing: the Shaker gene. *Trends Genet.* **4**, 297–299 (1988).
36. Schwarz, T. L., Tempel, B. L., Papazian, D. M., Jan, Y. N. & Jan, L. Y. Multiple potassium-channel components are produced by alternative splicing at the Shaker locus in *Drosophila*. *Nature* **331**, 137–142 (1988).
37. Littleton, J. T. & Ganetzky, B. Ion channels and synaptic organization: analysis of the *Drosophila* genome. *Neuron* **26**, 35–43 (2000).
38. Walcourt, A., Scott, R. L. & Nash, H. A. Blockage of one class of potassium channel alters the effectiveness of halothane in a brain circuit of *Drosophila*. *Anesth. Analg.* **92**, 535–541 (2001).
39. Rogero, O., Hammerle, B. & Tejedor, F. J. Diverse expression and distribution of Shaker potassium channels during the development of the *Drosophila* nervous system. *J. Neurosci.* **17**, 5108–5118 (1997).
40. Cuello, L. G., Cortes, D. M. & Perozo, E. Molecular architecture of the KvAP voltage-dependent K<sup>+</sup> channel in a lipid bilayer. *Science* **306**, 491–495 (2004).
41. Li-Smerin, Y., Hackos, D. H. & Swartz, K. J. Alpha-helical structural elements within the voltage-sensing domains of a K<sup>+</sup> channel. *J. Gen. Physiol.* **115**, 33–50 (2000).
42. Hendricks, J. C. *et al.* Gender dimorphism in the role of cycle (BMAL1) in rest, rest regulation, and longevity in *Drosophila melanogaster*. *J. Biol. Rhythms* **18**, 12–25 (2003).
43. Taheri, S. & Mignot, E. The genetics of sleep disorders. *Lancet Neurol.* **1**, 242–250 (2002).
44. Hendricks, J. C. *et al.* A non-circadian role for cAMP signaling and CREB activity in *Drosophila* rest homeostasis. *Nature Neurosci.* **4**, 1108–1115 (2001).
45. Benington, J. H., Woudenberg, M. C. & Heller, H. C. Apamin, a selective SK potassium channel blocker, suppresses REM sleep without a compensatory rebound. *Brain Res.* **692**, 86–92 (1995).
46. Vyazovskiy, V. V. *et al.* Sleep EEG in mice that are deficient in the potassium channel subunit K.v.3.2. *Brain Res.* **947**, 204–211 (2002).
47. Espinosa, F., Marks, G., Heintz, N. & Joho, R. H. Increased motor drive and sleep loss in mice lacking Kv3-type potassium channels. *Genes Brain Behav.* **3**, 90–100 (2004).
48. Liguori, R. *et al.* Morvan's syndrome: peripheral and central nervous system and cardiac involvement with antibodies to voltage-gated potassium channels. *Brain* **124**, 2417–2426 (2001).
49. Levine, J. D., Funes, P., Dowse, H. B. & Hall, J. C. Signal analysis of behavioral and molecular cycles. *BMC Neurosci.* **3**, 1 (2002).
50. Timpe, L. C. & Jan, L. Y. Gene dosage and complementation analysis of the Shaker locus in *Drosophila*. *J. Neurosci.* **7**, 1307–1317 (1987).

**Supplementary Information** accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work was supported by a grant from the United States Defense Advanced Research Projects Agency. B.G. is funded by NIH. We thank C. Holladay for technical assistance, M. Heisenberg and his laboratory for help with the heat box, and M. Rosbash for the circadian software.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to G.T. (gtononi@wisc.edu).

# Massive infection and loss of memory CD4<sup>+</sup> T cells in multiple tissues during acute SIV infection

Joseph J. Mattapallil<sup>1</sup>, Daniel C. Douek<sup>2</sup>, Brenna Hill<sup>2</sup>, Yoshiaki Nishimura<sup>3</sup>, Malcolm Martin<sup>3</sup> & Mario Roederer<sup>1</sup>

<sup>1</sup>ImmunoTechnology Section and <sup>2</sup>Human Immunology Section, Vaccine Research Center, and <sup>3</sup>Laboratory of Molecular Microbiology, NIAID, NIH, Bethesda, Maryland 20892, USA

It has recently been established that both acute human immunodeficiency virus (HIV) and simian immunodeficiency virus (SIV) infections are accompanied by a dramatic and selective loss of memory CD4<sup>+</sup> T cells predominantly from the mucosal surfaces. The mechanism underlying this depletion of memory CD4<sup>+</sup> T cells (that is, T-helper cells specific to previously encountered pathogens) has not been defined. Using highly sensitive, quantitative polymerase chain reaction together with precise sorting of different subsets of CD4<sup>+</sup> T cells in various tissues, we show that this loss is explained by a massive infection of memory CD4<sup>+</sup> T cells by the virus. Specifically, 30–60% of CD4<sup>+</sup> memory T cells throughout the body are infected by SIV at the peak of infection, and most of these infected cells disappear within four days. Furthermore, our data demonstrate that the depletion of memory CD4<sup>+</sup> T cells occurs to a similar extent in all tissues. As a consequence, over one-half of all memory CD4<sup>+</sup> T cells in SIV-infected macaques are destroyed directly by viral infection during the acute phase—an insult that certainly heralds subsequent immunodeficiency. Our findings point to the importance of reducing the cell-associated viral load during acute infection through therapeutic or vaccination strategies.

Chronic HIV infection is characterized by a steady but generally slow loss of CD4<sup>+</sup> T cells, of both naive and memory phenotypes. In contrast, the loss of CD4<sup>+</sup> T cells from peripheral blood during acute infection has been generally regarded to be modest and transient. However, studies using the SIV infection model have documented that acute infection is accompanied by a marked depletion of CD4<sup>+</sup> memory T cells primarily in mucosal tissues<sup>1,2</sup>; this has recently been confirmed in humans infected with HIV<sup>3</sup>. These studies suggest that mucosal CD4<sup>+</sup> T cells are the primary target for HIV infection and replication because of their high expression of the viral co-receptor CCR5 (ref. 4) as well as their relatively activated state<sup>5</sup>. Furthermore, it was observed that CD4<sup>+</sup> T cells with a mucosal homing phenotype were essentially completely absent, even from the peripheral blood<sup>6</sup>. As a consequence, the gut-associated mucosal tissue is viewed as the most important site of active viral replication and T-cell depletion during acute infection, owing to the concentration of viral targets at this site.

The frequency of infected CD4<sup>+</sup> T cells in the chronic phase of SIV/HIV infection is too low (0.01–1%) to account for this ongoing depletion by viral infection<sup>7–10</sup>. The mechanism for the massive and rapid loss during the acute phase is unknown. To answer this question, we longitudinally sampled blood, inguinal lymph nodes, as well as mucosal tissue and associated lymph nodes from the same animals before and after SIV infection at a regular and high frequency. In this way, we could address the issue of tissue distribution of T cells, a matter that clouds the interpretation of measurements taken solely from peripheral blood. Furthermore, we performed highly sensitive quantitative polymerase chain reaction (qPCR) analyses to determine which subsets of CD4<sup>+</sup> T cells were infected, and to what extent virus propagated through these subsets.

## T-cell dynamics

With the objective of delineating changes in T-cell subsets during early acute infection, we longitudinally evaluated the effects of SIV infection on naive and memory T-cell subsets in blood, inguinal and

mesenteric lymph nodes and jejunal mucosa, and compared them to pre-infection values (Figs 1 and 2). In nearly all animals (rhesus macaques, *Macaca mulatta*) there is an early increase in CD4<sup>+</sup> T cells at day 3 after infection in the blood (Fig. 2a). This is probably a re-distribution from other tissues, caused by an early, perhaps innate, immune response to the infection. By day 14 after infection, there is a 20% decrease in the number of CD4<sup>+</sup> T cells from the blood and lymph nodes. In contrast, as previously reported<sup>1,2,4</sup>, the jejunum exhibits a near total loss of CD4<sup>+</sup> T cells. Notably, these losses are significantly greater than the fraction of CCR5<sup>+</sup> CD4<sup>+</sup> T cells present in these tissues.

Because CCR5, the SIV co-receptor<sup>11,12</sup>, is expressed solely on memory CD4<sup>+</sup> T cells, we hypothesized that quantifying memory subset dynamics would provide a better view of the ongoing infection. As shown in Fig. 1a, naive T cells in these animals typically comprise 50–75% of T cells in blood (in adult humans this averages 50%). Consequently, the dynamics of CD4<sup>+</sup> memory T cells are far more marked than for the total compartment. An early expansion (day 3) is followed by the loss of 60–80% of memory T cells. This loss is seen in multiple tissues, although it appears first in the blood and lymph nodes and a few days later in the jejunum<sup>1,2</sup>. Thus, the SIV-induced depletion of CD4<sup>+</sup> T cells during acute infection is not principally restricted to T cells of mucosal origin. However, in terms of the total number of CD4<sup>+</sup> T cells lost from an animal, most are from the mucosa, as the greatest number of T cells are resident there.

We quantified the CCR5-expressing T cells in the various tissues. As previously described<sup>4,6</sup>, these cells disappear quickly from the peripheral blood—beginning before the onset of viraemia. Interestingly, even CCR5<sup>+</sup> CD8<sup>+</sup> T cells were found to disappear from the blood; however, unlike CCR5<sup>+</sup> CD4<sup>+</sup> T cells, they continued to be found in other tissues.

## Viral dynamics

In order to understand the role of viral infection of CD4<sup>+</sup> T cells on their dynamics, we quantified both plasma and cell-associated virus (Fig. 3). We used qPCR<sup>7</sup> on bulk-sorted subsets of T cells to

determine the extent to which direct SIV infection could account for the loss of memory CD4<sup>+</sup> T cells. Starting shortly after infection, the number of SIV-Gag copies in memory CD4<sup>+</sup> T-cell subsets rose steadily, peaking at day 10 after infection in all tissues examined (Fig. 3b). In fact, at peak, we observed a very high number of SIV-Gag copies (50–200,000 per 10<sup>5</sup> CD4<sup>+</sup> memory T cells). The level of SIV-Gag copies then declined dramatically by day 14 after infection (Fig. 3b). These kinetics reflect the highly dynamic nature of the infection, and underscore the importance of measuring parameters of infection at the infection peak (day 10–11 after infection) as well as after the peak.

The amount of SIV-Gag DNA in memory CD4<sup>+</sup> T cells correlated well with the plasma viral loads (Fig. 3c), similar to what has been shown for the chronic phase of HIV infection in humans<sup>7</sup>. This suggests that plasma virus arises directly from infected memory CD4<sup>+</sup> T cells.

The high levels of SIV-Gag DNA could be due to a very high number of copies in a small number of cells, or to a low number of copies in nearly all memory cells. To distinguish between these possibilities, we quantified the amount of SIV-Gag DNA in single-sorted memory T cells (Table 1). We calibrated our results using a cell line carrying a single copy of proviral DNA. We thus determined that CD4<sup>+</sup> memory T cells in all the tissues examined carried, on average, 1.5 copies of Gag DNA (Table 1). This measurement is in good agreement with published data for HIV infection in humans, where infected CD4<sup>+</sup> T cells in lymph nodes carried on average two copies of viral DNA<sup>13,14</sup>. From the quantification of Gag DNA at the single cell level, we could determine that the fraction of cells infected in the different tissues (Table 1) ranged from 30 to 60% of all

memory CD4<sup>+</sup> T cells at the peak of infection.

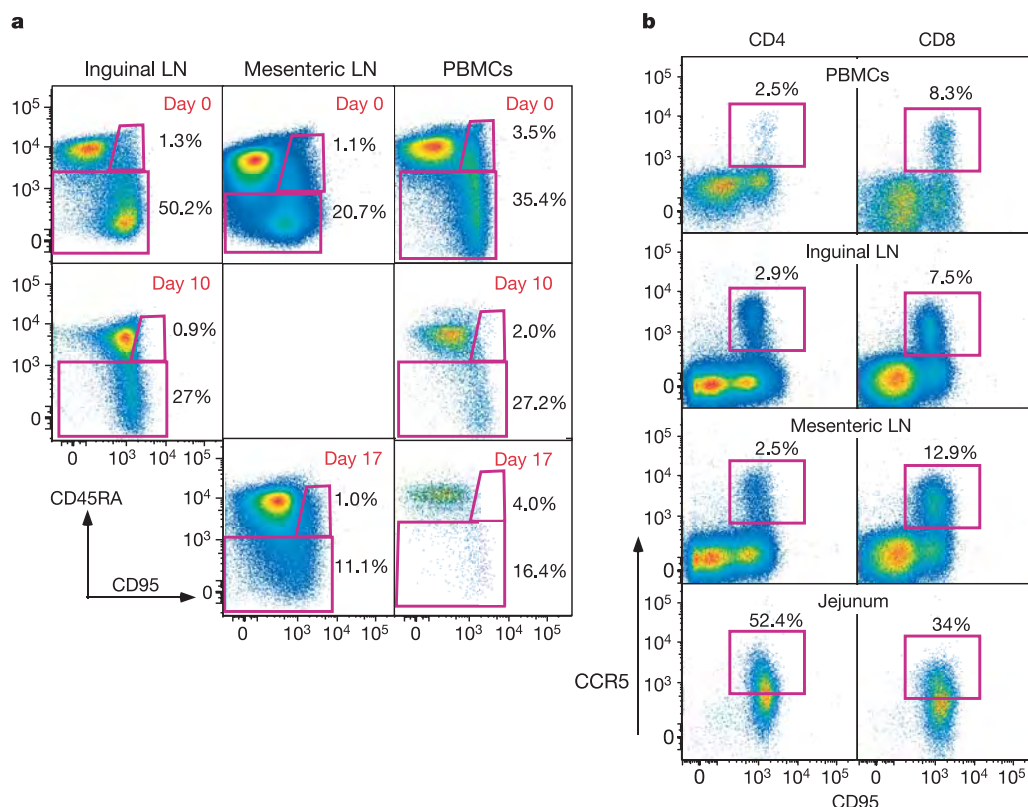
Quantification of viral load in purified naive T cells confirmed the intrinsic resistance of these cells to SIV infection<sup>15</sup>. On the basis of qPCR of 168 wells of single- or low-number cell sorts, we estimate that less than 0.5% (1 out of 200) of naive T cells were infected by SIV even at day 10 after infection (data not shown).

By day 14, the cell-associated viral load in all tissues dropped about 80% from peak values. Because this viral load is quantified by the presence of Gag DNA, the loss corresponds to the loss of 80% of infected CD4<sup>+</sup> T cells. Because we found similar kinetics in all tissues, we do not believe that a redistribution of cells accounts for this loss.

### Apparent CCR5 expression does not correlate with infection

SIV uses CCR5 as its co-receptor<sup>11,12</sup>; the lack of expression of CCR5 by naive T cells presumably underlies their resistance to SIV infection. However, the infection rate in the tissues described above (30–60%) far exceeds the well-described expression of CCR5 by memory T cells in these tissues as well as the rate that we found for our animals (~3% of total CD4<sup>+</sup> T cells in peripheral blood and lymph nodes, ~50% in mucosa; Fig. 1).

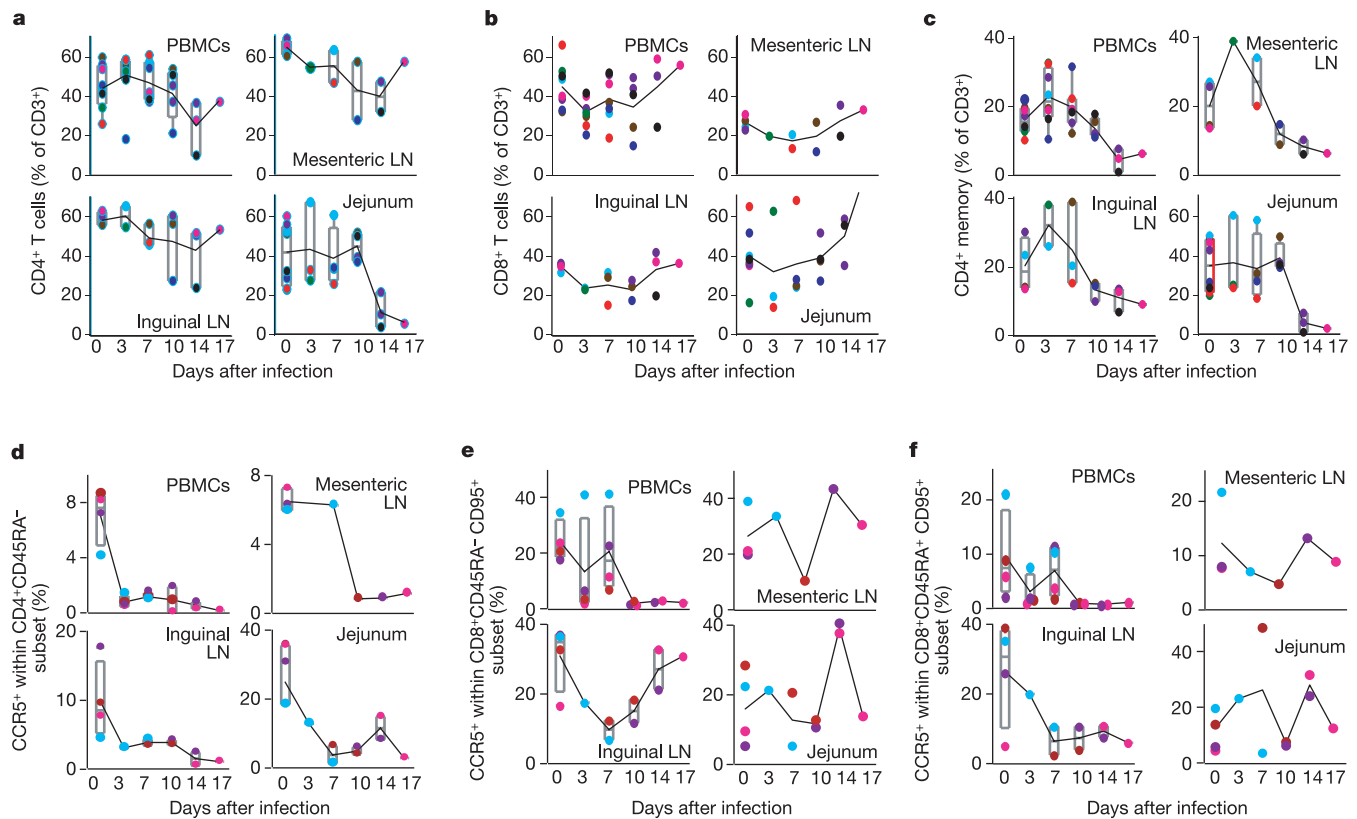
We determined that the ability to measure CCR5 expression by flow cytometry does not correlate with the presence of SIV in memory CD4<sup>+</sup> T cells. The level of viral infection in sorted CCR5<sup>+</sup> cells was essentially the same as that for sorted CCR5<sup>−</sup> cells (Fig. 3b). We hypothesized that CCR5<sup>−</sup> memory T cells are not truly negative for CCR5 expression—they simply do not express enough receptor to be detectable by flow cytometry. To address this, we quantified CCR5 messenger RNA expression on sorted naive, CCR5<sup>−</sup> and



**Figure 1** Identification of subsets of CD4<sup>+</sup> and CD8<sup>+</sup> T cells in a non-human primate (rhesus macaque). **a**, Naive CD4<sup>+</sup> T cells (CD45RA<sup>+</sup>CD95<sup>−</sup>) and two subsets of memory CD4<sup>+</sup> T cells (CD45RA<sup>−</sup>, or CD45RA<sup>+</sup>CD95<sup>+</sup>) can be identified in tissues of rhesus macaques<sup>6,20</sup>. Before infection, most cells in the blood and lymph nodes are of the naive

phenotype; however, acute SIV infection is accompanied by a selective loss of the memory subsets. LN, lymph node. **b**, CCR5 expression is restricted to memory T cells in both CD4<sup>+</sup> and CD8<sup>+</sup> T cells.





**Figure 2** Dynamics of CD4<sup>+</sup> and CD8<sup>+</sup> subsets during acute SIV infection. **a, b**, Modest declines in CD4<sup>+</sup> T cells (**a**) occur in all tissues except the jejunum, where a near total loss occurs between days 10 and 14; CD8<sup>+</sup> T cells (**b**) show the opposite trend. Nearly all jejunal T cells are of the memory phenotype (Fig. 1b); in other tissues naive T cells predominate. **c**, After an early expansion of CD4<sup>+</sup> memory T cells (CD45RA<sup>+</sup> and CD45RA<sup>-</sup> subsets), most are lost in the subsequent week. y-axis values in **a–c** are a percentage of total T cells. **d**, CCR5<sup>+</sup> CD4<sup>+</sup> T cells show early changes, before the onset of detectable viraemia. The percentages are as a fraction of CD45RA<sup>-</sup> memory T cells;

the fraction within total CD4<sup>+</sup> T cells is two- to fourfold lower (Fig. 1b). **e, f**, CCR5<sup>+</sup> CD8<sup>+</sup> T cells also show changes in their response to the SIV challenge, for both the CD45RA<sup>-</sup> (**e**) and CD45RA<sup>+</sup> (**f**) memory subsets, including a selective loss from blood. These changes illustrate the capacity of CCR5<sup>+</sup> T cells to redistribute in the body, limiting interpretations of the loss of these cells from measurements solely on blood. Boxes represent interquartile range, where applicable. Within this figure, colours identify individual animals.

CCR5<sup>+</sup> memory cells (Fig. 4; sorted cells were >99% pure). CCR5<sup>-</sup> memory cells had ~20-fold more mRNA than naive T cells; similarly they had ~20-fold less than CCR5<sup>+</sup> memory T cells. Thus, many ‘CCR5<sup>-</sup>’ memory T cells may in fact express sufficient levels of CCR5 to render them susceptible to SIV infection—as was shown *in vitro* for HIV infection<sup>16</sup>. An alternative explanation is that CCR5 expression is highly labile and does not preclude the possibility that infected cells were CCR5<sup>+</sup> at the time they were infected. In any case, CCR5 expression on CD4<sup>+</sup> memory T cells, measured at a single time point, cannot be used to describe viral dynamics accurately.

## Discussion

We find that acute SIV infection is accompanied by a surprisingly high rate of infection and subsequent deletion of memory CD4<sup>+</sup> T cells, across a wide spectrum of tissues in the host. Previous estimates of the infection rate of CD4<sup>+</sup> T cells were well under 1% (refs 7–10). The markedly higher levels of infection that occur during acute viraemia last only a few days: after the peak, at which 30–60% of all memory CD4<sup>+</sup> T cells are infected, the cell-associated viral loads drop very rapidly. Indeed, during the first 4 days after peak, we found that about 80% of infected cells are eliminated, either by virus-induced cytolysis, or by an immune-

**Table 1** Quantification of SIV infection at the single cell level

| Tissue                 | Gag copies per cell* | Infection rate (%)† |                   |                    |
|------------------------|----------------------|---------------------|-------------------|--------------------|
|                        |                      | 10 cells per well   | 30 cells per well | 100 cells per well |
| PBMCs                  | 1.45                 | 29                  | 30                | 31                 |
| Mesenteric lymph nodes | 1.51                 | 57                  | 50                | 55                 |
| Jejunum                | 1.61                 | 59                  | 60                | 62                 |
| Inguinal lymph nodes   | 1.45                 | 44                  | 44                | 59                 |

\* For each tissue (day 10 after infection), single CD4<sup>+</sup> memory T cells were deposited in 72 wells by a flow cytometer. The average signal from the positive wells was converted to Gag copies per cell based on a calibration using a cell line with a single integrated virion. The efficiency of recovering PCR signal from the cell line was 77%.

† A total of 12–36 wells, each containing 10, 30, or 100 cells deposited by a flow cytometer, were quantified for SIV-Gag DNA; the resulting signal was divided by the average number of copies per cell (second column) to define the fraction of infected cells. This percentage correlates very well with the values obtained from the bulk qPCR analyses (see Fig. 3), for which the cell number was calculated from albumin DNA qPCR.

mediated mechanism. Because 30–60% of all memory CD4<sup>+</sup> T cells were infected, this corresponds to an elimination of 24–48% of all memory T cells between day 10 and day 14. Loss of memory CD4<sup>+</sup> T cells continues after day 14, albeit to a lesser extent.

The loss of infected cells measured by qPCR (Fig. 3b) corresponds with the loss of memory CD4<sup>+</sup> T cells measured phenotypically (Fig. 2c). Thus, we can explain the loss of CD4<sup>+</sup> memory T cells during acute infection solely by mechanisms related to the direct infection of these cells by SIV: either by SIV-induced cytolysis or an SIV-specific immune response that kills the infected cells. No ‘bystander’ killing mechanisms need be invoked to account for the observed T-cell dynamics, although we cannot rule out that such mechanisms still occur.

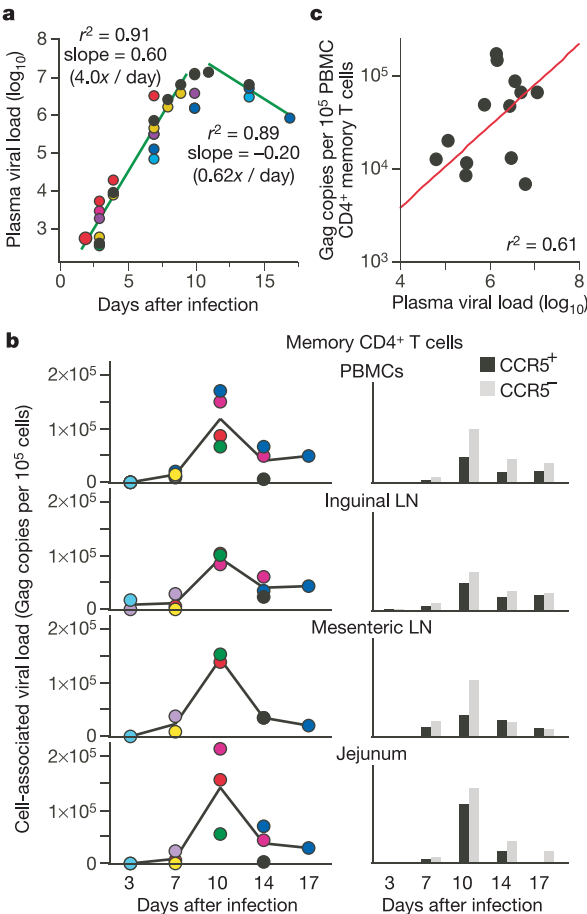
The rapidity and extent with which peak viraemia is achieved and resolved makes detailed measurement of viral and T-cell dynamics particularly difficult. In order to derive the best measurements for quantitative kinetic analysis, nearly daily sampling may be necessary. In addition, when modelling phenotype-defined dynamics, it is important to consider that there are many changes occurring in peripheral blood well before the onset of viraemia (Fig. 2), suggesting that the acute innate

immune response is affecting lymphocyte trafficking. Such dynamics have not been considered in the modelling of the acute infection and may markedly affect the interpretations of such models.

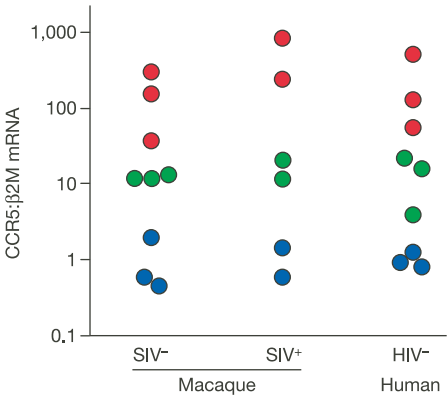
The fact that most memory T cells can be infected is not unexpected: lentiviruses are known to infect non-cycling cells, and resting memory T cells have been shown to harbour SIV<sup>17,18</sup>. However, the efficient depletion of infected cells at days 10–14 means that the infection of these cells must be accompanied by at least a minimal expression of viral genes. The selective depletion of the infected cells is due to either CD8-mediated killing (requiring sufficient expression of viral genes to render the infected cells ‘visible’ to cytotoxic T lymphocytes) or a viral-induced cytopathic event involving some level of cellular activation. In any case, our observations demonstrate that most memory CD4<sup>+</sup> T cells can support a sufficiently productive infection, leading to their demise.

The extent to which memory CD4<sup>+</sup> T cells could be infected during the acute stage was a surprise. Why is it that the fraction of infected cells at the chronic stage is so low (<<1%; refs 7–10)? Clearly the immune response is dramatically affecting the balance in the chronic phase, and perhaps the viruses produced at later stages contain a much greater proportion of non-infectious particles. It is remarkable that, even at late stages, in the presence of relatively high amounts of virus (10<sup>4</sup> virions per ml), only a small proportion of the susceptible memory CD4<sup>+</sup> T-cell population carries virus.

We found that naive CD4<sup>+</sup> T cells are highly (or completely) resistant to SIV infection (Fig. 3b), in agreement with a previous study<sup>15</sup>. As a consequence, the dynamics of peripheral blood CD4<sup>+</sup> T cells is highly tempered because of the presence of a large (and variable) fraction of these resistant cells: the massive depletion of the memory compartment is masked by the majority of the population (naive T cells) that is unaffected by SIV. We recommend that future



**Figure 3** Viral dynamics during acute infection. **a**, Plasma virus was first detected at day 3 after infection and peaked at day 10, at 10<sup>7</sup> copies per ml. Note that animals were killed at various times throughout the time course. A least-squares (log-linear) regression of the rise and fall in viral loads is shown. **b**, Cell-associated viral loads for sorted memory CD4<sup>+</sup> T cells. The peak in cell-associated viral load coincided with the plasma viral load (day 10 after infection), and fell as much as 80% by day 14. CCR5<sup>+</sup> and CCR5<sup>-</sup> memory CD4<sup>+</sup> T cells were sorted from these tissues; the average viral load for the two subsets is shown in the bar charts. **c**, The plasma viral load is reasonably well correlated with the memory CD4<sup>+</sup> T-cell-associated viral load.



**Figure 4** Expression of CCR5 mRNA by T cells. Quantitative RT-PCR was used to compare the CCR5 mRNA levels in sorted CD4<sup>+</sup> naive (blue), CCR5<sup>+</sup> memory (red) and CCR5<sup>-</sup> memory (green) T cells in three SIV-negative macaques, two SIV-positive macaques (at day 10 after infection) and three HIV-negative humans.

| Table 2 Tissue sampling schedule |                        |               |                            |                              |
|----------------------------------|------------------------|---------------|----------------------------|------------------------------|
| Number of animals                | Blood (day)            | Jejunum (day) | Inguinal lymph nodes (day) | Mesenteric lymph nodes (day) |
| 1                                | P, 0, 3                | P, 0, 3       | P, 0, 3                    | P, 3                         |
| 2                                | P, 0, 3, 7             | P, 3, 7       | P, 3, 7                    | P, 7                         |
| 2                                | P, 0, 3, 7, 10         | P, 7, 10      | P, 7, 10                   | P, 10                        |
| 2                                | P, 0, 3, 7, 10, 14     | P, 10, 14     | P, 10, 14                  | P, 14                        |
| 1                                | P, 0, 3, 7, 10, 14, 17 | P, 14, 17     | P, 14, 17                  | P, 17                        |

Listed are the days after infection (where P indicates 1 month before infection) at which samples were collected from each animal in the group. The last collection date was a necropsy; all animals sustained only a single survival surgery.

studies of SIV T-cell dynamics (and, for example, evaluation of vaccine efficacy) be tailored to longitudinally distinguish naive and memory CD4<sup>+</sup> T cells—the animal-to-animal variation in the naive/memory balance could mask clinically important changes in the memory CD4<sup>+</sup> T cells.

Acute SIV infection is considerably more marked than has been previously described. There is an extensive loss of memory CD4<sup>+</sup> T cells not only from mucosal tissues (where most CD4<sup>+</sup> T cells reside) but also from organized lymph nodes and peripheral blood. The high rate of infection of these T cells is a sufficient mechanism to account for their loss during acute infection; no bystander mechanisms need be invoked. It is clear that a central aim for effective vaccines (or early intervention drug therapy) must be to prevent this massive destruction of the CD4<sup>+</sup> memory compartment by tempering the cell-associated viral load at the peak time point. □

## Methods

### Animals, infection and samples

Eight colony-bred healthy rhesus macaques (*Macaca mulatta*) housed at Bioqual Inc were used in this study. Animals were housed in accordance with American Association for Accreditation of Laboratory Animal Care guidelines and were sero-negative for SIV, simian retrovirus and simian T-cell leukaemia virus type-1. Animals were infected with 100 animal infectious doses of uncloned pathogenic SIVmac251 intravenously (courtesy of N. Letvin); plasma and tissue samples were collected at various time points by biopsy or necropsy (Table 2).

### Tissue sampling

Peripheral blood mononuclear cells (PBMCs) were isolated by density gradient centrifugation. Jejunal biopsy and necropsy samples were isolated as previously described<sup>2</sup>. Lymph node cell suspensions were made by pushing tissue through a small-gauge syringe needle. Plasma viral RNA levels were determined by real-time PCR (ABI Prism 7700 sequence detection system, Applied Biosystems) using reverse-transcribed viral RNA as templates, as described<sup>19</sup>.

### Antibodies and flow cytometry

All antibodies were purchased from PharMingen, either conjugated or unconjugated and derivatized in our laboratory. All reagents were validated and titrated using rhesus macaque PBMCs. For phenotypic analysis, freshly isolated cells were labelled simultaneously with the following combinations of antibodies: CD3-Cy7-allophycocyanin (APC), CD8-Cy5.5-phycoerythrin (PE), CD4-cascade blue, CD45RA-TRPE, CD95-APC, CCR5-PE. Labelled cells were fixed with 0.5% paraformaldehyde and analysed using a modified Becton Dickinson Digital Vantage. At least one million total events were collected. To determine which subsets supported viral infection, naive and memory CD4<sup>+</sup> T cells (discriminated on the basis of CD45RA and CD95 expression<sup>20</sup>) were sorted into tubes and subjected to qPCR assay for measuring SIV-Gag DNA<sup>7</sup>. To determine whether these T-cell subsets harboured single or multiple copies of SIV-Gag DNA, cells were sorted directly into 96-well PCR plates at varying frequencies of 1–100 cells per well and subjected to qPCR analysis. For each sort, a cell line containing a single copy of proviral SIV DNA (described below) was also sorted to validate the qPCR assay.

### qPCR assay for SIV-Gag DNA

T-cell-associated viral DNA was measured by a quantitative PCR assay for SIV gag using a Perkin-Elmer ABI 7700 instrument as previously described<sup>7</sup>, and using SIV gag primers and probe as described previously<sup>21</sup>.

### Quantification of CCR5 mRNA

T-cell subsets were sorted directly into RNeasy Lysate (Ambion Inc.) and centrifuged at 10,000g for 3 min. Supernatants were discarded and total RNA was extracted using RNeasy Lysate-4-PCR and treated with 2 units of DNase for 30 min at 37 °C (Ambion Inc.). After DNase treatment, mRNA was purified using the Oligotex mRNA extraction kit as per the manufacturer's instructions (Qiagen).

Purified mRNA was added directly to a one-step quantitative RT-PCR reaction containing Superscript RT-Platinum Taq enzyme mix (Invitrogen). CCR5 and normalizing gene probes (β2-microglobulin, or β2M) were labelled with a 5' FAM reporter and 3' BHQ1 quencher (Biosource). We used the following oligonucleotide sequences: CCR5 forward primer, GTCCCTTCTGGGCTCACTAT; CCR5 reverse primer, CCCTGTCAAGAGTTGACACATTGTA; CCR5 probe, FAM-TCCAAAGTCCCACTGGGAGCAG-BHQ1; β2M forward primer, GCTGGCGCTACTCTCTCTTCT; β2M reverse primer, GGATGGCGTGAGTAAACCTGAA; β2M probe, FAM-CCTGGAGGCTATCCAGCGTACTCCAAAG-BHQ1. Expression levels of human and non-human primate CCR5 were normalized to β2-microglobulin and calculated based on the ΔΔC<sub>T</sub> method<sup>22</sup>; Fig. 4 shows the values normalized to the average value for naive CD4<sup>+</sup> T cells in each group.

### Generation of a cell line with a single copy of SIV proviral DNA

CEM x174 cells were infected with SIVmac316 (ref. 23) at a multiplicity of infection of

0.01. As previous studies had shown that cultured cells surviving acute infection harbour low copy numbers of integrated and defective proviral DNA<sup>24</sup>, cells were collected 8 weeks after inoculation. Single cells were sorted into 96-well round-bottom plates; a total of 76 clones containing proviral DNA was obtained from 384 individual wells. Southern blotting of restricted genomic DNA was performed to determine the copy number(s) of SIV proviral DNA in each cell, as previously reported<sup>25</sup>. Three clones (2A7, 2E10 and 3D8), each containing a single copy of SIV DNA, were identified; clone 3D8 was used in this study. This clone produces no detectable virus.

### Data analysis

Flow cytometric data were analysed using FlowJo version 6.1 (Tree Star, Inc.). Statistical analyses were computed with JMP (SAS Institute).

Received 22 December 2004; accepted 25 February 2005; doi:10.1038/nature03501.

Published online 27 March 2005.

1. Veazey, R. S. *et al.* Gastrointestinal tract as a major site of CD4<sup>+</sup> T cell depletion and viral replication in SIV infection. *Science* **280**, 427–431 (1998).
2. Mattapallil, J. J., Smit-McBride, Z., McChesney, M. & Dandekar, S. Intestinal intraepithelial lymphocytes are primed for γ-interferon and MIP-1β expression and display antiviral cytotoxic activity despite severe CD4<sup>+</sup> T-cell depletion in primary simian immunodeficiency virus infection. *J. Virol.* **72**, 6421–6429 (1998).
3. Brenchley, J. M. *et al.* CD4<sup>+</sup> T cell depletion during all stages of HIV disease occurs predominantly in the gastrointestinal tract. *J. Exp. Med.* **200**, 749–759 (2004).
4. Veazey, R. S. *et al.* Dynamics of CCR5 expression by CD4<sup>+</sup> T cells in lymphoid tissues during simian immunodeficiency virus infection. *J. Virol.* **74**, 11001–11007 (2000).
5. Veazey, R. S. *et al.* Identifying the target cell in primary simian immunodeficiency virus (SIV) infection: highly activated memory CD4<sup>+</sup> T cells are rapidly eliminated in early SIV infection *in vivo*. *J. Virol.* **74**, 57–64 (2000).
6. Mattapallil, J. J., Letvin, N. L. & Roederer, M. T-cell dynamics during acute SIV infection. *AIDS* **18**, 13–23 (2004).
7. Douek, D. C. *et al.* HIV preferentially infects HIV-specific CD4<sup>+</sup> T cells. *Nature* **417**, 95–98 (2002).
8. Haase, A. T. *et al.* Quantitative image analysis of HIV-1 infection in lymphoid tissue. *Science* **274**, 985–989 (1996).
9. Haase, A. T. Population biology of HIV-1 infection: viral and CD4<sup>+</sup> T cell demographics and dynamics in lymphatic tissues. *Annu. Rev. Immunol.* **17**, 625–656 (1999).
10. Lassen, K., Han, Y., Zhou, Y., Siliciano, J. & Siliciano, R. F. The multifactorial nature of HIV-1 latency. *Trends Mol. Med.* **10**, 525–531 (2004).
11. Chen, Z., Zhou, P., Ho, D. D., Landau, N. R. & Marx, P. A. Genetically divergent strains of simian immunodeficiency virus use CCR5 as a coreceptor for entry. *J. Virol.* **71**, 2705–2714 (1997).
12. Marcon, L. *et al.* Utilization of C–C chemokine receptor 5 by the envelope glycoproteins of a pathogenic simian immunodeficiency virus, SIVmac239. *J. Virol.* **71**, 2522–2527 (1997).
13. Jung, A. *et al.* Multiply infected spleen cells in HIV patients. *Nature* **418**, 144 (2002).
14. Gratton, S., Cheynier, R., Dumaourier, M. J., Oksenhendler, E. & Wain-Hobson, S. Highly restricted spread of HIV-1 and multiply infected cells within splenic germinal centers. *Proc. Natl Acad. Sci. USA* **97**, 14566–14571 (2000).
15. Nishimura, Y. *et al.* Highly pathogenic SHIVs and SIVs target different CD4<sup>+</sup> T cell subsets in rhesus monkeys, explaining their divergent clinical courses. *Proc. Natl Acad. Sci. USA* **101**, 12324–12329 (2004).
16. Mengozzi, M. *et al.* Naive CD4 T cells inhibit CD28-costimulated R5-tropic HIV replication in memory CD4 T cells. *Proc. Natl Acad. Sci. USA* **98**, 11644–11649 (2001).
17. Zhang, Z. *et al.* Sexual transmission and propagation of SIV and HIV in resting and activated CD4<sup>+</sup> T cells. *Science* **286**, 1353–1357 (1999).
18. Zhang, Z. Q. *et al.* Roles of substrate availability and infection of resting and activated CD4<sup>+</sup> T cells in transmission and acute simian immunodeficiency virus infection. *Proc. Natl Acad. Sci. USA* **101**, 5640–5645 (2004).
19. Endo, Y. *et al.* Short- and long-term clinical outcomes in rhesus monkeys inoculated with a highly pathogenic chimeric simian/human immunodeficiency virus. *J. Virol.* **74**, 6935–6945 (2000).
20. Pitcher, C. J. *et al.* Development and homeostasis of T cell memory in rhesus macaque. *J. Immunol.* **168**, 29–43 (2002).
21. Lifson, J. D. *et al.* Role of CD8<sup>+</sup> lymphocytes in control of simian immunodeficiency virus infection and resistance to challenge after transient early antiretroviral treatment. *J. Virol.* **75**, 10187–10199 (2001).
22. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the ΔΔC<sub>T</sub> Method. *Methods* **25**, 402–408 (2001).
23. Mori, K., Ringle, D. J., Kodama, T. & Desrosiers, R. C. Complex determinants of macrophage tropism in env of simian immunodeficiency virus. *J. Virol.* **66**, 2067–2075 (1992).
24. Folks, T. M. *et al.* Biological and biochemical characterization of a cloned Leu-3- cell surviving infection with the acquired immune deficiency syndrome retrovirus. *J. Exp. Med.* **164**, 280–290 (1986).
25. Folks, T. *et al.* Characterization of a continuous T-cell line susceptible to the cytopathic effects of the acquired immunodeficiency syndrome (AIDS)-associated retrovirus. *Proc. Natl Acad. Sci. USA* **82**, 4539–4543 (1985).

**Acknowledgements** We thank R. Koup for critical comments on the manuscript; S. Peretto, J. Yu, R. Nguyen, D. Ambrozak and other members of the VRC Laboratory of Immunology for advice and technical help; M. St Claire for assistance with the animals; and S. Rao, V. Dang and J.-P. Todd for their assistance during the course of this study.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to M.R. (Roederer@nih.gov).



# An exceptionally bright flare from SGR 1806–20 and the origins of short-duration $\gamma$ -ray bursts

K. Hurley<sup>1</sup>, S. E. Boggs<sup>1,2</sup>, D. M. Smith<sup>3</sup>, R. C. Duncan<sup>4</sup>, R. Lin<sup>1</sup>, A. Zoglauer<sup>1</sup>, S. Krucker<sup>1</sup>, G. Hurford<sup>1</sup>, H. Hudson<sup>1</sup>, C. Wigger<sup>5</sup>, W. Hajdas<sup>5</sup>, C. Thompson<sup>6</sup>, I. Mitrofanov<sup>7</sup>, A. Sanin<sup>7</sup>, W. Boynton<sup>8</sup>, C. Fellows<sup>8</sup>, A. von Kienlin<sup>9</sup>, G. Lichti<sup>9</sup>, A. Rau<sup>9</sup> & T. Cline<sup>10</sup>

<sup>1</sup>UC Berkeley Space Sciences Laboratory, Berkeley, California 94720-7450, USA

<sup>2</sup>University of California, Department of Physics, Berkeley, California 94720, USA

<sup>3</sup>Physics Department and Santa Cruz Institute for Particle Physics, University of California, Santa Cruz, California 95064, USA

<sup>4</sup>University of Texas, Department of Astronomy, Austin, Texas 78712, USA

<sup>5</sup>Paul Scherrer Institute, 5232 Villigen PSI, Switzerland

<sup>6</sup>Canadian Institute of Theoretical Astrophysics, 60 St George Street, Toronto, Ontario M5S 3H8, Canada

<sup>7</sup>Space Research Institute (IKI), GSP7, Moscow 117997, Russia

<sup>8</sup>University of Arizona, Department of Planetary Sciences, Tucson, Arizona 85721, USA

<sup>9</sup>Max-Planck-Institut für extraterrestrische Physik, Giessenbachstrasse (Postfach 1312), 85748 (85741) Garching, Germany

<sup>10</sup>NASA Goddard Space Flight Center, Code 661, Greenbelt, Maryland 20771, USA

**Soft- $\gamma$ -ray repeaters (SGRs) are galactic X-ray stars that emit numerous short-duration (about 0.1 s) bursts of hard X-rays during sporadic active periods. They are thought to be magnetars: strongly magnetized neutron stars with emissions powered by the dissipation of magnetic energy. Here we report the detection of a long (380 s) giant flare from SGR 1806–20, which was much more luminous than any previous transient event observed in our Galaxy. (In the first 0.2 s, the flare released as much energy as the Sun radiates in a quarter of a million years.) Its power can be explained by a catastrophic instability involving global crust failure and magnetic reconnection on a magnetar, with possible large-scale untwisting of magnetic field lines outside the star. From a great distance this event would appear to be a short-duration, hard-spectrum cosmic  $\gamma$ -ray burst. At least a significant fraction of the mysterious short-duration  $\gamma$ -ray bursts may therefore come from extragalactic magnetars.**

In the magnetar model, SGRs are isolated neutron stars with teragauss exterior magnetic fields<sup>1–4</sup> and even stronger fields within<sup>5,6</sup>, making them the most strongly-magnetized objects in the Universe. Four SGRs are known. Three of them have now emitted giant flares<sup>7,8</sup>. These exceptionally energetic outbursts begin with a brief ( $\sim 0.2$  s) spike of  $\gamma$ -rays with energies up to several MeV, containing most of the flare energy. The spikes are followed by tails lasting minutes, during which hard-X-ray emissions gradually fade while oscillating at the rotation period of the neutron star.

The first-known giant flare, observed on 5 March 1979, came from SGR 0525–66 in the Large Magellanic Cloud. Its fluence implied an energy  $\geq 6 \times 10^{44}$  erg (ref. 9). The second-known giant flare came from an SGR in our Galaxy, SGR 1900+14, on 27 August 1998. Its energy, in hard X-rays and  $\gamma$ -rays, was  $\sim 2 \times 10^{44}$  erg (refs 8, 10). Here we describe a third giant flare, which came from the galactic SGR 1806–20 on 27 December 2004. Particle and  $\gamma$ -ray detectors onboard the Reuven Ramaty High Energy Solar Spectroscopic Imager (RHESSI), and particle detectors aboard the Wind spacecraft, indicate that this event was  $\sim 100$  times more energetic than the 27 August flare. Its initial  $\gamma$ -ray spike had a quasi-blackbody spectrum, characteristic of a relativistic pair/photon outflow with an energetically small contamination of baryons. This is consistent with the catastrophic release of (nearly) pure magnetic energy from a magnetar<sup>3</sup>. The tremendous luminosity of the initial spike means that similar events could be detected from distant galaxies. This could account for some, and perhaps all, of the mysterious short-duration, hard-spectrum cosmic  $\gamma$ -ray bursts (GRBs).

## The giant flare from SGR 1806–20

On 27 December 2004, the International Gamma-Ray Astrophysics

Laboratory<sup>11</sup> (INTEGRAL) reported the detection of a spectacular flare. Four other missions in the third interplanetary network of GRB detectors (the High Energy Neutron Detector and Gamma Sensor Head aboard Mars Odyssey<sup>12</sup>, the solar-pointing RHESSI<sup>13</sup>, particle and  $\gamma$ -ray detectors aboard the Wind spacecraft<sup>14</sup>, and NASA's recently launched GRB observatory Swift<sup>15</sup>) also reported this event. The light curve is shown in Fig. 1. Triangulation constrains the flare position to a portion of an annulus consistent with SGR 1806–20's position (annulus centre J2000, right ascension 15 h 56 m 37 s, declination  $-20^\circ 13' 50''$ , annulus radius  $30.887 \pm 0.030^\circ$ ). No other known or candidate SGR lies within this area of the sky. SGR 1806–20 was  $5.25^\circ$  from the Sun at the time of these observations.

A  $\sim 1$ -s-long precursor was observed 142 s before the flare, with a roughly flat-topped profile (Fig. 1 inset). Its spectrum can be fitted with an optically thin thermal bremsstrahlung function with  $kT \approx 15$  keV. The precursor's  $>3$ -keV fluence was  $1.8 \times 10^{-4}$  erg cm<sup>-2</sup>, implying an energy of  $4.8 \times 10^{42} d_{15}^2$  erg, where  $d_{15} = (d/15 \text{ kpc})$ , and  $d$  is the distance to SGR 1806–20. Note that  $0.8 < d_{15} < 1$  is likely for SGR 1806–20, owing to the apparent association of the SGR with a compact ( $\sim 10$  arcsec) stellar cluster<sup>16,17</sup>. The large energy and unusual light curve of the precursor distinguish it from most common SGR bursts. This and its proximity in time to the giant flare suggest that it is causally related.

The initial spike of the giant flare lasted for  $\sim 0.2$  s. Its rise and fall times were  $\tau_{\text{rise}} \leq 1$  ms and  $\tau_{\text{decay}} \approx 65$  ms, similar to those of the other giant flares<sup>8,18</sup>. The spike's intensity drove all X- and  $\gamma$ -ray detectors into saturation, but particle detectors aboard RHESSI and Wind made reliable measurements. (The Supplementary Information describes our extensive Monte Carlo simulations of these particle detectors and has a full discussion of systematic

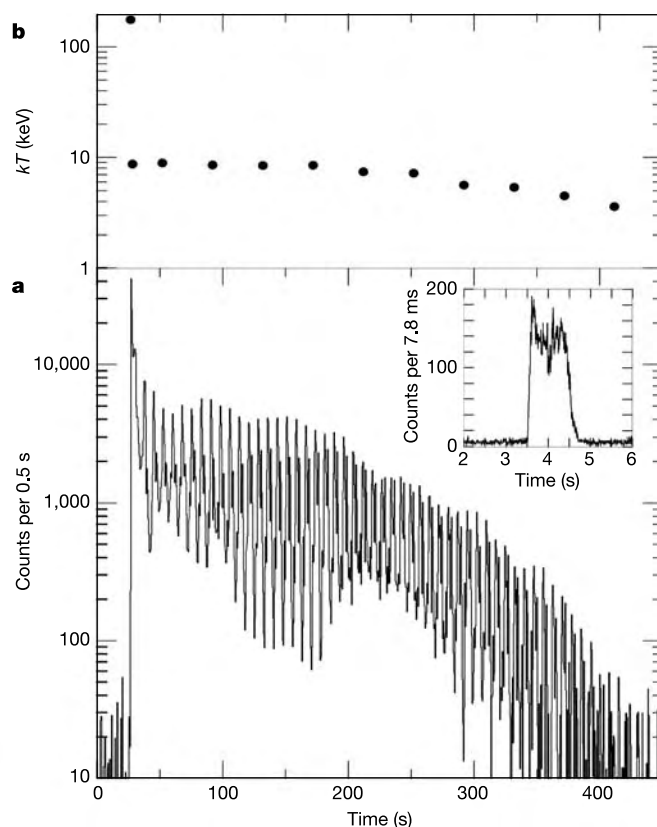
uncertainties.) The RHESSI particle detector data imply a spike fluence in photons  $>30$  keV of  $(1.36 \pm 0.35) \text{ erg cm}^{-2}$ , making this the most intense cosmic or solar transient ever observed (in terms of photon energy flux at Earth). The time-resolved energy spectrum, as measured by the Wind particle detectors, is consistent with a cooling blackbody (Fig. 2) with average temperature  $T_{\text{spike}} = (175 \pm 25) \text{ keV}$ . The spike energy is thus  $E_{\text{spike}} = (3.7 \pm 0.9) \times 10^{46} d_{15}^2 \text{ erg}$ , assuming isotropic emission. The peak flux in the first 0.125 s was  $L_{\text{spike}} = 2 \times 10^{47} d_{15}^2 \text{ erg s}^{-1}$ . Evidently, this event briefly outshone all the stars in the Galaxy put together by a factor of  $\sim 10^3$ .

The spike was followed by a hard-X-ray tail modulated with a period of 7.56 s, detected by the RHESSI  $\gamma$ -ray detectors, which were by this time unsaturated, for 380 s. This period agrees with the neutron star rotation period as inferred from cyclic modulations of its quiescent soft-X-ray counterpart<sup>2</sup>. The fluence in 3–100-keV

photons during the tail phase is  $4.6 \times 10^{-3} \text{ erg cm}^{-2}$  or  $E_{\text{tail}} \approx 1.2 \times 10^{44} d_{15}^2 \text{ erg}$ .

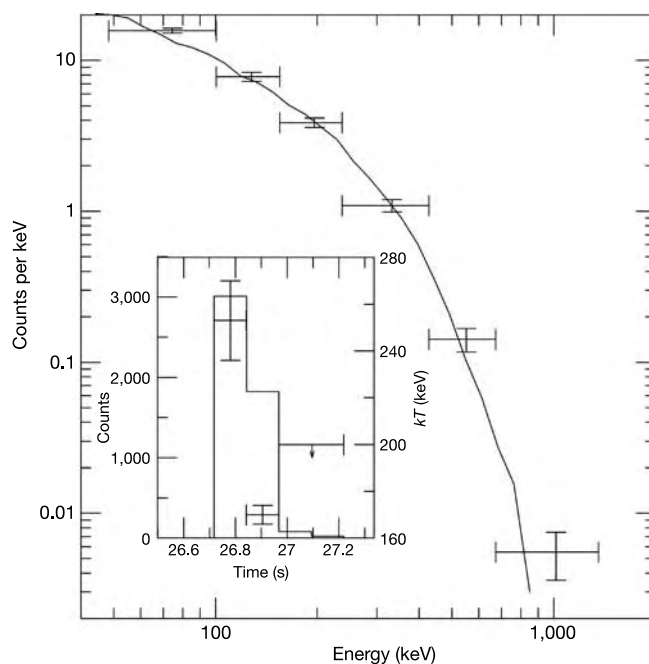
### Physical interpretation

This event can be understood as a result of a catastrophic instability in a magnetar. Strong shearing of the neutron star's magnetic field, combined with growing thermal pressure, appears to have forced an opening of the field outward, launching a hot fireball. The release of energy above a rate of  $\sim 10^{42} \text{ erg s}^{-1}$  (less than one part in  $10^4$  of the peak flare luminosity) into the magnetosphere leads to the formation of a hot, thermal pair plasma ( $kT \approx 0.1\text{--}1 \text{ MeV}$ )<sup>19</sup>. The fast initial rise  $\tau_{\text{rise}} \leq 1 \text{ ms}$  is consistent with a magnetospheric instability with characteristic time  $\tau_{\text{mag}} \approx (R/0.1V_A) \approx 0.3 \text{ ms}$ , where  $R \approx 10 \text{ km}$  and  $V_A \approx c$  is the Alfvén velocity in the magnetosphere, and  $c$  is the speed of light<sup>3</sup>. This process must have occurred repeatedly, given that the hard initial spike persisted for a duration  $\sim 10^3 \tau_{\text{mag}}$ . Indeed, there is evidence for spike variability in this and other giant flares<sup>8,20,21</sup>. The resulting outflow emitted a quasi-blackbody spectrum as it became optically thin, with spectral temperature comparable to the temperature at its base, because declining temperature in the outflow is compensated by the relativistic blue-shift<sup>22</sup>. For luminosity  $L_{\text{spike}} = 10^{47} L_{47} \text{ erg s}^{-1}$ , where  $L_{47} = L/10^{47} \text{ erg s}^{-1}$  and  $L$  is the luminosity emerging from a zone with radius  $R \approx 10 \text{ km}$ , the expected spectral temperature is  $T_{\text{spike}} = (L_{\text{spike}}/4\pi acR^2)^{0.25} = 200 L_{47}^{0.25} \text{ keV}$ , neglecting complications of magnetospheric stresses and intermittency. Almost all the pairs annihilated, and the outflow was only weakly polluted by baryons, as is clear from the extended, weak radio afterglow that followed the flare<sup>23,53</sup>. Note that we do not expect strong beaming of such powerful emissions from such a slowly rotating star.



**Figure 1** Profiles of the 27 December 2004 giant flare. **a**, 20–100-keV time history plotted with 0.5-s resolution, from the RHESSI  $\gamma$ -ray detectors. Zero seconds corresponds to 77,400 s Universal Time (UT). In this plot, the flare began with the spike at 26.64 s and saturated the detectors within 1 ms. The detectors emerged from saturation on the falling edge 200 ms later and remained unsaturated after that. Photons with energies  $\geq 20$  keV are unattenuated; thus the amplitude variations in the oscillatory phase are real, and are not caused by any known instrumental effect (Supplementary Information). Inset, time history of the precursor with 8-ms resolution. Zero corresponds to 77,280 s UT.

**b**, Spectral temperature versus time. The temperature of the spike was determined by the RHESSI and Wind particle detectors; the temperatures of the oscillatory phase were measured by the RHESSI  $\gamma$ -ray detectors. Although RHESSI measured time- and energy-tagged photons  $>3$  keV continuously, unattenuated spectra were measured for short ‘snapshot’ intervals only twice in each 4.06-s spacecraft spin period during the oscillatory phase (Supplementary Information). Preliminary spectral analysis (3–100 keV), using the RHESSI on-axis response matrices, are generally consistent with a single-temperature blackbody or optically thin thermal bremsstrahlung model; the blackbody temperatures have been plotted. The formal uncertainties in the oscillatory phase are smaller than the data points and are not shown.



**Figure 2** Spectrum and time history of the initial spike, from the RHESSI and Wind particle detectors. The crosses show the spectrum measured by the Wind 3D O detector<sup>52</sup> with coarse time resolution that averages over the peak. The error bars are  $1\sigma$ , plus 10% systematic errors. The line is the best-fitting blackbody convolved with the detector response function; its temperature is  $175 \pm 25 \text{ keV}$  (Supplementary Information). Inset, the time history of the peak (histogram, left-hand scale) and of the blackbody temperature (error bars, right-hand scale) with 0.125-s resolution, from the RHESSI particle detector (ref. 35 and Supplementary Information). The error bars are  $1\sigma$ , plus 25% systematic errors.

When the outflow ceased, a trapped fireball was evidently left behind: an optically thick photon-pair plasma confined by closed field lines near the star<sup>3,17</sup>. The luminosity and lifetime of the tail (see the fitted curve in Fig. 3) are consistent with a fireball cooling rate that is limited by the transparency of the surface layers, where the temperature is  $\sim 10$  keV and the plasma is dominated by ions and electrons<sup>3,19,24</sup>. The condition that the magnetic field must be strong enough to confine energy  $E_{\text{tail}}$  within a distance  $\Delta R \approx 10$  km of the star yields a rough bound on the dipole field,  $B_{\text{dipole}} > 2 \times 10^{14} (\Delta R/10 \text{ km})^{-3/2} [(1 + \Delta R/R)/2]^3 \text{ G}$ , similar to bounds implied by the previous giant flares<sup>3,8</sup>.

A clue to the nature of the instability comes from the spike's  $\sim 0.2$ -s duration, which is similar to the durations of other giant flare spikes<sup>7,8,18</sup> and of most other SGR bursts<sup>25</sup>. In the magnetar model, SGR activity results from the unwinding of a strong, toroidal magnetic field inside the star, and the transfer of magnetic helicity across the surface<sup>19,26</sup>. Such a twist propagates along the poloidal magnetic field  $B_p = 10^{15} B_{p15} \text{ G}$  with a speed  $V_A \approx B_p / (4\pi\rho)^{-0.5}$  that is weakly dependent on the twist amplitude. The time to cross the neutron star interior (density  $\rho 10^{15} \rho_{15} \text{ g cm}^{-3}$ ) is  $\Delta t \approx 2R/V_A \approx 0.2 B_{p15}^{-1} \text{ s}$ .

Thus the 27 December event could have been a crustal instability that drove helicity from the star<sup>19,26</sup>. The unwinding of a toroidal magnetic field embedded in the crust is strongly impeded by the stable stratification and near-incompressibility of the crust<sup>19</sup>. Because of the energetic cost of forming isolated dislocation surfaces that cross the magnetic flux surfaces, the crust must undergo smooth and vertically differential torsional motion when it fails, which requires a fundamental breakdown of its solid structure. The

maximum field energy which can be released is estimated by balancing elastic and magnetic stresses in the crust:  $E_{\text{max}} \approx 1 \times 10^{46} (\theta_{\text{max}}/10^{-2})^2 B_{p15}^{-2} \text{ erg}$ , where  $\theta_{\text{max}}$  is the yield strain. Supplying the energy of the 27 December flare thus requires a relatively large yield strain, as well as a large twist of the crust with angular displacement approaching  $\sim 0.5 B_{p15}^{-1}$  radian.

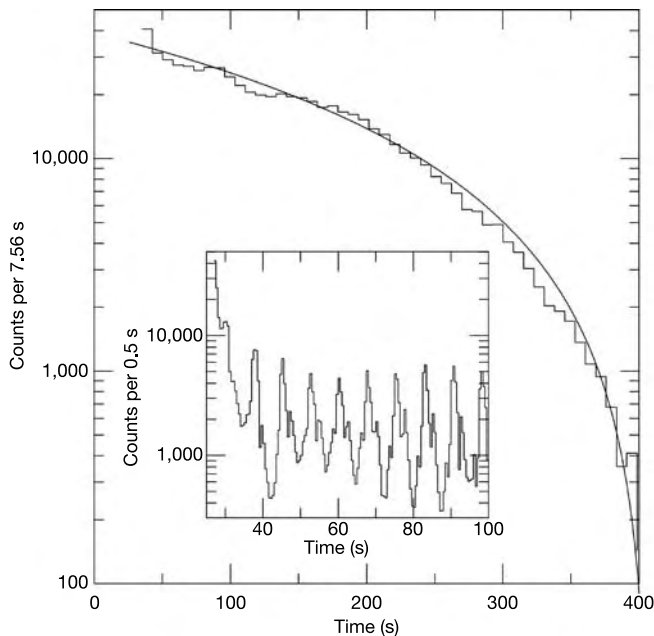
Since March 2004, SGR 1806–20 has been very burst-active<sup>27</sup>, while its quiescent X-ray brightness has increased by a factor of 2 to 3, and its spectrum has hardened dramatically<sup>28</sup>. Evidently, crust failure has enhanced the twist in the external magnetic field, with growing magnetospheric currents<sup>26</sup>. The free energy of such an exterior magnetic twist can reach a modest fraction ( $\sim 10^{-1}$ ) of the untwisted exterior dipole field energy,  $E_{\text{twist}} \approx 10^{-2} B_{\text{dip}}^2 R^3 \approx 10^{46} B_{p15}^2 \text{ erg}$ , with more energy in the non-potential components of higher multipoles. Some of this energy could be catastrophically released via reconnective simplification of the magnetosphere<sup>26,29</sup>. An extreme possibility, consistent with the flare energy, is a global magnetospheric untwisting. This would predict a dramatic post-flare drop in the stellar spin-down rate, as well as greatly diminished, softened and less strongly pulsed X-ray emissions. However, a pure magnetospheric instability would proceed much faster than  $\sim 0.2$  s. Note also that the detection of accelerated spin-down<sup>30</sup> several months after previous active periods of SGRs 1806–20 and 1900+14 betrays a net increase in the magnetospheric twist during the X-ray bursts, and in the 27 August 1998 giant flare. Observations of SGR 1806–20's spin-down over the coming year will provide important constraints on the location of the non-potential magnetic field that was dissipated during the flare.

### Short-duration GRBs and magnetars

If observed from a great distance, only the brief, initial hard spike of the 27 December flare would be evident. Thus distant extragalactic magnetar flares (MFs) would resemble the mysterious short-duration GRBs<sup>31,32</sup>. These hard-spectrum events have long been recognized as a separate class of GRBs<sup>33–37</sup> but have never been identified with any counterparts<sup>38</sup>.

The Burst and Transient Source Experiment (BATSE) on the Compton Gamma-Ray Observatory was a landmark experiment of the 1990s that produced a catalogue<sup>39</sup> of more than 2,000 GRBs. How many of these bursts were MFs? Taking the 27 December event as our prototype and adopting the 50% trigger-efficiency flux<sup>40</sup> of  $0.5 \text{ photons cm}^{-2} \text{ s}^{-1}$  for the 256-ms timescale yields a BATSE sampling depth of  $D_{\text{BATSE}} = 30 d_{15} \text{ Mpc}$ . If such events generally happen once every  $\tau = 30 \text{ yr}$  in galaxies like the Milky Way (such as has now occurred in the Milky Way itself) then the BATSE detection rate of MFs is  $\dot{N}_{\text{BATSE}} = 19 d_{15}^3 (\tau/30 \text{ yr})^{-1} \text{ yr}^{-1}$ . Here we have estimated the effective number of galaxies like the Milky Way within  $D_{\text{BATSE}}$  of Earth by multiplying the local blue luminosity density<sup>41</sup>  $j_b = 5.8 \times 10^{41} h_{70} \text{ erg Mpc}^{-3}$  by the sampling volume  $(4\pi/3) D_{\text{BATSE}}^3$ , and dividing by the blue luminosity of the Milky Way as estimated in the Supplementary Information. We use blue emissions as a benchmark because SGRs are Population I objects, the post-supernova remnants of massive, short-lived, blue stars. Thus, over 9.5 yr of operation with half-sky coverage, BATSE probably detected  $180 d_{15}^3 (\tau/30 \text{ yr})^{-1}$  MFs, representing  $0.4 d_{15}^3 (\tau/30 \text{ yr})^{-1}$  of all BATSE short-duration bursts. There is evidence of 100-s-long soft tails in the co-added time histories of many short-duration BATSE GRBs<sup>42,43</sup>, but not in any single event. For the brightest observed BATSE short-duration, hard-spectrum GRB (trigger number 6293), we find that the ratio of the tail-to-peak fluence is  $< 0.5\%$ , compared to our measured ratio for the 27 December event of 0.34%. Thus BATSE was not sensitive enough to have detected MF tails in single bursts.

The GRB observatory Swift<sup>44</sup> was designed, in part, to unravel the short-duration GRB mystery. How many MFs will Swift spot? The Swift Burst Alert Telescope has a photon flux sensitivity (50–300 keV) that is  $\sim 5$  times better than BATSE<sup>45</sup>, corresponding

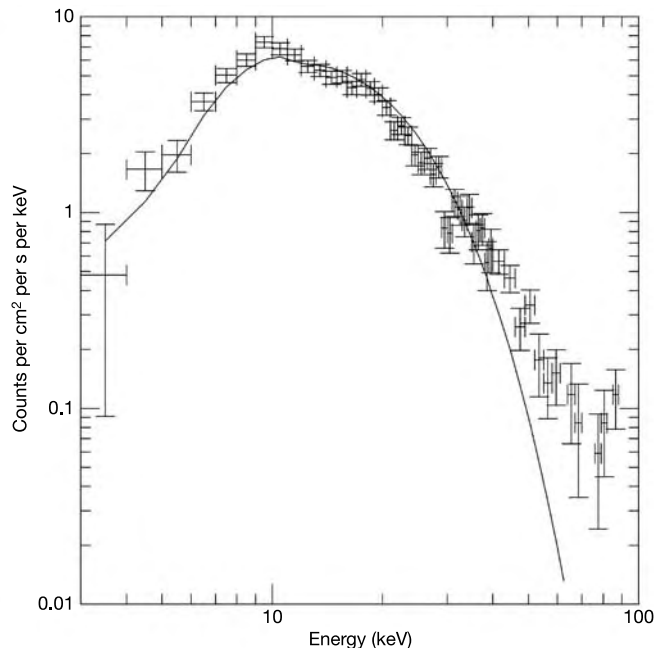


**Figure 3** Time-averaged counts in the tail phase of the giant flare, compared with the ‘trapped fireball’ model. Zero corresponds to 77,280 s UT. The step plot shows the RHESSI  $\gamma$ -ray detector data averaged over the 7.56-s rotation period of the neutron star. It is fitted by a simple model (smooth curve) that describes the emission from the cool surface of a magnetically confined plasma as it contracts and evaporates in a finite time:  $L_x(t) = L_0 [1 - (t/t_{\text{evap}})]^{a/(1-a)}$  (ref. 49). We find  $t_{\text{evap}} = 382 \pm 3 \text{ s}$ , and the index  $a = 0.606 \pm 0.003$  is near the value  $a = 2/3$  expected for a homogeneous, spherical trapped fireball<sup>19,49</sup>. Inset, RHESSI  $\gamma$ -ray detector light curve for the first ten cycles of the flare tail. The energy range is 20–100 keV. The first peak of the trapped fireball emission is evident on the falling edge of the hard spike at  $t = 30 \text{ s}$ . A changing two-peaked pulse-interpulse structure is present.



to a trigger threshold of  $\sim 0.10$  photons  $\text{cm}^{-2} \text{s}^{-1}$ . Thus for our prototype MF,  $D_{\text{Swift}} = 70d_{15}$  Mpc. The expected rate of MF detections, given Swift's sky coverage of 1.4 steradians, is then  $\dot{N}_{\text{Swift}} = 53d_{15}^3(\tau/30\text{yr})^{-1} \text{yr}^{-1}$ , or about one MF per week. Of course, the galactic rate of MFs,  $\Gamma = \tau^{-1}$ , is very uncertain. Given that there has occurred one MF with peak luminosity in the range  $10^{47} \text{erg s}^{-1}$  in our Galaxy during  $t_0 = 30$  yr of observations, the bayesian probability distribution for the underlying galactic rate  $\Gamma$  of such bright MFs is  $(dP/d\Gamma) = t_0 \exp(-\Gamma t_0)$ , with expected value  $\langle \Gamma \rangle = t_0^{-1}$ . This implies that the probability that Swift will detect one or more MF per month is 80% for  $d_{15} = 1$ . The probabilities of detecting one or more event per {3, 6, 12, 24} months are {93, 96, 98, 99}%, respectively. Even if  $d = 10$  kpc, the probabilities would be {78, 88, 94, 97}%. The prospects for observing MFs during Swift's 24-month prime mission are excellent.

Of course, all of the above estimates idealize MFs as 'standard candles' defined by the 27 December prototype. The actual luminosity function of MFs is unknown. It is possible that some MFs are significantly brighter than the 27 December event. For example, a magnetic instability on a rare magnetar with  $B_{\text{dipole}} \approx 10^{16}$  G could release  $10^{48}$  erg, and be detected by Swift out to  $\sim 1$  Gpc. Nevertheless, we suspect that MFs constitute only a substantial subset of BATSE Class II GRBs, not all of them. The 175-keV blackbody spectrum would probably result in a significantly higher hardness ratio than that of the average short-duration burst<sup>37</sup>. The fact that Class II GRBs have  $\langle V/V_{\text{max}} \rangle < 0.5$  does not seem consistent with all these events being local. Moreover, no galaxies at  $D < 100$  Mpc were found for the Interplanetary Network positions of four short-duration GRBs<sup>38</sup>.

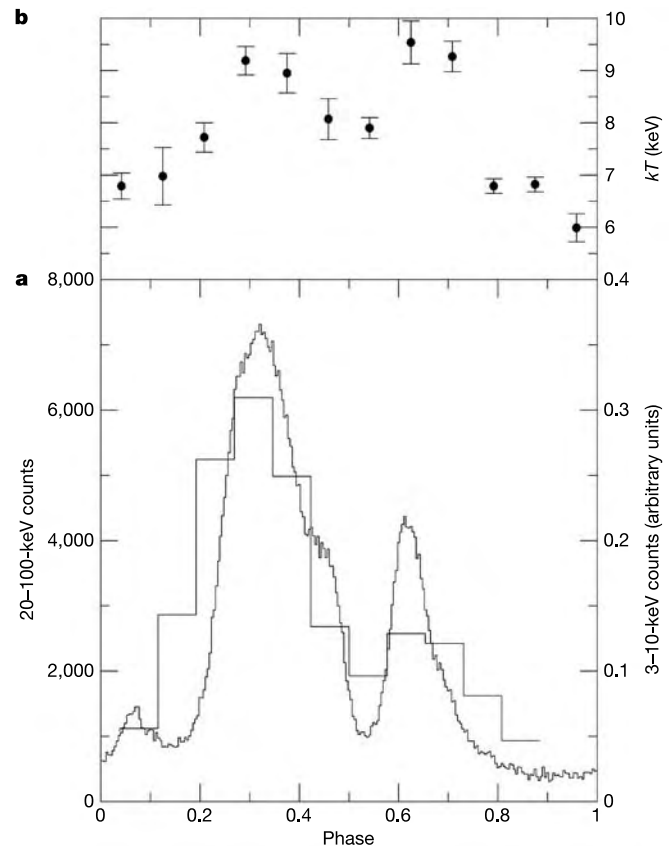


**Figure 4** 3–100-keV phase-averaged energy spectrum of the pulsed tail, from the RHESSI  $\gamma$ -ray detectors. The crosses show the measured spectrum with  $1\sigma$  statistical error bars; the solid line represents a fit to a blackbody function  $E^2(\exp(E/kT) - 1)^{-1}$ , where  $E$  is the energy and  $kT = 5.1 \pm 1.0$  keV. This spectrum is averaged over various phases between 272 and 400 s in Fig. 1, corresponding to intervals where the photons could reach the detectors passing through a minimum amount of intervening materials (Supplementary Information). An optically thin thermal bremsstrahlung function with  $kT \approx 22$  keV also provides a reasonable fit. The spectra show evidence of deviations from both models, probably due to the use of an approximate response matrix<sup>24</sup>.

## Studying extragalactic magnetars

Swift can identify MFs via their positional correlations with galaxies, allowing the source distances from Earth to be inferred. A spiral galaxy of size  $\sim 30$  kpc at distance  $D_{\text{Swift}}$  spans  $\sim 3.4$  arcmin, comparable to the Swift BAT location accuracy of  $\Delta\theta_{\text{BAT}} \approx 1$ –4 arcmin. This localization can be greatly improved, to an accuracy of  $\lesssim 10$  arcsec, if the oscillating tail of the flare is detected by Swift's X-ray Telescope (XRT) when it slews to observe the burst site within about 1 min. Our measurements of soft X-ray emissions in the giant flare tail (Fig. 4) make it possible to assess the prospects of XRT acquisition for the first time. Extrapolating our X-ray spectral fits down to 0.3 keV, we find that the 27 December pulsating tail produced a 0.3–10-keV incident fluence of  $(0.18$ – $1.6) \times 10^{-3} \text{erg cm}^{-2}$ . The threshold fluence for XRT detection<sup>44</sup> is  $2 \times 10^{-10} \text{erg cm}^{-2}$ , so that the 27 December flare tail could be marginally detected to a distance of  $D_{\text{tail}} = (10$ – $40)d_{15}$  Mpc. Thus only the nearest fraction  $(D_{\text{tail}}/D_{\text{Swift}})^3 \approx 0.2$  of all MFs spotted by Swift will have detectable tails. We have verified that the soft X-rays are strongly pulsed (Fig. 5). For events within about 8 Mpc, simulations indicate that the magnetar's rotation period can be reliably determined. For more distant sources, the spectrum and the rapid flux decay will distinguish magnetar tail emissions from cosmic GRB afterglows.

The prospects of detecting extragalactic MFs with the Swift Ultra-Violet and Optical Telescope (UVOT) or ground-based optical telescopes are not wholly bleak. The trapped fireball is too tiny to



**Figure 5** Detailed profiles of the oscillations, from the RHESSI  $\gamma$ -ray detectors. **a**, RHESSI light curve for the oscillatory portion of the giant flare, folded modulo the 7.56-s neutron star rotation period (20–100 keV, fine resolution curve, and 3–10 keV, coarse resolution curve). **b**, The blackbody spectral temperature  $kT$ . The radius of the emitting surface varies between  $\sim 18$  and 40 km at 15 kpc. The error bars represent  $1\sigma$  statistical uncertainties.

emit detectably in this waveband. However, we can scale directly from the observed radio afterglow<sup>23</sup>, which had spectral index  $\alpha = -0.7$  and time decay  $t^{-1.5}$  in the optically thin regime. Extrapolating to  $10^{14.5}$  Hz gives  $L_{\text{opt}} \approx 4 \times 10^{37} t_3^{-1.5} \text{ erg s}^{-1}$  at a time  $10^3 t_3$  s post-flare. Such a source would have a brightness of 22nd magnitude at 1 Mpc for  $t_3 \approx 1$ .

Prospects are even better for the detection of X-ray afterglows<sup>32</sup>. SGR 1900+14 emitted strong nonthermal X-rays in the aftermath of the 27 August 1998 event<sup>46</sup>, thought to be due to a heated magnetar crust<sup>47</sup>. If afterglow energy scales linearly with flare energy, as found in less energetic events<sup>48</sup>, then a MF like the 27 December event would glow brighter by a factor of  $f \approx 100$ , suggesting  $L_X \approx 2 \times 10^{39} (f/100) (t/1 \text{ h})^{-0.7} \text{ erg s}^{-1}$ . This could be detected by the Chandra X-ray Telescope within  $D_{\text{Chandra}} \approx 30 (f/100)^{0.5} (\Delta t_{\text{obs}}/10^4 \text{ s})^{0.5} (t/10 \text{ h})^{-0.35} \text{ Mpc}$  in an observation of duration  $\Delta t_{\text{obs}} \ll t$ .

## New horizons and speculations

The detection of extragalactic magnetars, if achieved by Swift, will open up a new field of astronomy. A catalogue of giant flare spikes, once assembled, will contain a wealth of information about magnetic instabilities in neutron stars. Information about the luminosity function of MFs, their range of durations, and possible spectral diversity (suggested by measurements of the 27 August event<sup>8,49</sup>; note that less compact flows than that of the 27 December event could show nonthermal spectra) will constrain magnetar physics and population diversity. Unusually bright flares may be detected from very young magnetars with shorter rotation periods and stronger fields than are observed in galactic SGRs. (The birth-rate of SGRs is evidently low enough that no stars younger than  $\sim 10^3 - 10^4$  yr are observed in our galaxy.) MFs from very young magnetars may be disproportionately common in extragalactic studies because of their greater brightness and higher flare rate. More frequent cataclysms are expected in younger magnetars because magnetic diffusion slows down as stars age and cool<sup>6</sup>.

We emphasize that most SGR activity is ultimately powered by the strong toroidal interior field of a magnetar,  $B_\phi$ , which is the remnant of the rapid differential rotation which the neutron star experienced at birth<sup>1,5</sup>. The energy of this field,  $E_\phi \approx (1/6) B_\phi^2 R^3 \approx 2 \times 10^{49} B_{\phi 16}^2 \text{ erg}$ , where  $B_{\phi 16} = (B_\phi/10^{16} \text{ G})$ , can power many flares of  $\sim 10^{46} \text{ erg}$  over a star's lifetime. Magnetic helicity is gradually transported outward via ambipolar diffusion and Hall drift<sup>6</sup>, winding up the field within the crust and outside the star, and leading to catastrophic instabilities<sup>19,26</sup>. (Note, however, that the strong, internal toroidal field stabilizes a magnetar against catastrophic decay of the exterior dipole field; compare with refs 5 and 32.) Measurements of SGR 1806–20's spin-down over the coming year will reveal whether the exterior magnetic helicity increased or decreased during the 27 December event. SGR 1806–20 may come to resemble an anomalous X-ray pulsar, with a diminished spin-down rate and a softer X-ray spectrum. SGR 0526–66 developed these characteristics, indicating weakened magnetospheric currents, after the giant flare of 5 March 1979 (ref. 50). Sporadic, short bursts were observed from SGR 0525–66 until 1983 (ref. 51), but the source has not been observed to burst since then, suggesting that sub-crust stresses were (at least temporarily) relieved in the giant flare. We speculate that SGR 0526–66 and now SGR 1806–20 may have entered the 'low' phase in a magnetar activity cycle that involves changes in the rate of expulsion of magnetic helicity out of the star. □

Received 7 February; accepted 4 March 2005; doi:10.1038/nature03519.

- Duncan, R. & Thompson, C. Formation of very strongly magnetized neutron stars: implications for gamma-ray bursts. *Astrophys. J.* **392**, L9–L13 (1992).
- Kouveliotou, C. *et al.* An X-ray pulsar with a superstrong magnetic field in the soft gamma-ray repeater SGR 1806–20. *Nature* **393**, 235–237 (1998).
- Thompson, C. & Duncan, R. The soft gamma repeaters as very strongly magnetized neutron stars. I. Radiative mechanism for outbursts. *Mon. Not. R. Astron. Soc.* **275**, 255–300 (1995).

- Kouveliotou, C. *et al.* Discovery of a magnetar associated with the soft gamma repeater SGR 1900+14. *Astrophys. J.* **510**, L115–L118 (1999).
- Thompson, C. & Duncan, R. Neutron star dynamos and the origins of pulsar magnetism. *Astrophys. J.* **408**, 194–224 (1993).
- Thompson, C. & Duncan, R. C. The soft gamma repeaters as very strongly magnetized neutron stars II. Quiescent neutrino, X-ray and Alfvén wave emission. *Astrophys. J.* **473**, 322–342 (1996).
- Mazets, E. *et al.* A flaring X-ray pulsar in Dorado. *Nature* **282**, 587–589 (1979).
- Hurley, K. *et al.* A giant periodic flare from the soft  $\gamma$ -ray repeater SGR 1900+14. *Nature* **397**, 41–43 (1999).
- Evans, W. *et al.* Location of the gamma-ray transient event of 1979 March 5. *Astrophys. J.* **237**, L7–L9 (1980).
- Vrba, F. *et al.* The discovery of an embedded cluster of high-mass stars near SGR 1900+14. *Astrophys. J.* **533**, L17–L20 (2000).
- Borkowski, J. *et al.* Giant flare from SGR 1806–20 detected by INTEGRAL. *GCN Circ.* 2920 (2004).
- Hurley, K. *et al.* IPN localization of the giant flare from SGR 1806–20. *GCN Circ.* 2921 (2004).
- Boggs, S. *et al.* SGR 1806–20, RHESSI observations of the 041227 giant flare. *GCN Circ.* 2936 (2004).
- Mazets, E. *et al.* The giant outburst from SGR 1806–20. *GCN Circ.* 2922 (2004).
- Palmer, D. E. A giant  $\gamma$ -ray flare from the magnetar SGR 1806–20. *Nature* doi:10.1038/nature03525 (this issue).
- Corbel, S. & Eikenberry, S. The connection between W31, SGR 1806–20, and LBV 1806–20: Distance, extinction, and structure. *Astron. Astrophys.* **419**, 191–201 (2004).
- Figer, D. E., Najarro, F. & Kudritzki, R. P. The double-lined spectrum of LBV 1806–20. *Astrophys. J.* **610**, L109–L113 (2004).
- Cline, T. *et al.* Detection of a fast, intense and unusual gamma ray transient. *Astrophys. J.* **237**, L1–L5 (1980).
- Thompson, C. & Duncan, R. The giant flare of 1998 August 27 from SGR 1900+14. Radiative mechanism and physical constraints on the source. *Astrophys. J.* **561**, 980–1005 (2001).
- Terasawa, T. *et al.* Repeated injections of energy in the first 600 ms of the giant flare of SGR 1806–20. *Nature* doi:10.1038/nature03573 (this issue); preprint at <http://arXiv.org/astro-ph/0502315> (2004).
- Barat, C. *et al.* Fine time structure in the 1979 March 5 gamma ray burst. *Astron. Astrophys.* **126**, 400–402 (1983).
- Paczynski, B. Gamma-ray bursters at cosmological distances. *Astrophys. J.* **308**, L43–L46 (1986).
- Gaensler, B. M. *et al.* An expanding radio nebula produced by a giant flare from the magnetar SGR 1806–20. *Nature* doi:10.1038/nature03498 (this issue).
- Boggs, S. *et al.* The giant flare of December 27, 2004 from SGR 1806–20. *Astrophys. J.* (submitted).
- Gogus, E. *et al.* Temporal and spectral characteristics of short bursts from the soft gamma repeaters 1806–20 and 1900+14. *Astrophys. J.* **558**, 228–236 (2001).
- Thompson, C., Lyutikov, M. & Kulkarni, S. Electrodynamics of magnetars: implications for the persistent X-ray emission and spin-down of the soft gamma repeaters and anomalous X-ray pulsars. *Astrophys. J.* **574**, 332–355 (2002).
- Golenetskii, S. *et al.* Bright bursts from SGR 1806–20. *GCN Circ.* 2823 (2004).
- Woods, P. *et al.* Gradual brightening of SGR 1806–20. *Astronomer's Telegram* 313 (2004).
- Lyutikov, M. Explosive reconnection in magnetars. *Mon. Not. R. Astron. Soc.* **346**, 540–554 (1998).
- Woods, P. M. *et al.* Large torque variations in two soft gamma repeaters. *Astrophys. J.* **576**, 381–390 (2002).
- Duncan, R. Gamma-ray bursts from extragalactic magnetar flares. *AIP Conf. Proc.* **586**, 495–501 (eds Martel, H. & Wheeler, J. C.) (AIP, New York, 2001).
- Eichler, D. Waiting for the big one: a new class of soft gamma repeater outbursts. *Mon. Not. R. Astron. Soc.* **576**, 381–392 (2002).
- Mazets, E., Golenetskii, S., Guryan, Yu. & Ilyinskii, V. The 5 March 1979 event and the distinct class of short gamma bursts—are they of the same origin? *Astrophys. Space Sci.* **84**, 173–189 (1982).
- Dezalay, J.-P. *et al.* The hardness-duration diagram of gamma-ray bursts. *Astrophys. J.* **471**, L27–L30 (1996).
- Norris, J., Cline, T., Desai, U. & Teegarden, B. Frequency of fast, narrow gamma ray bursts. *Nature* **308**, 434–435 (1984).
- Hurley, K. Gamma-ray burst observations: past and future, in gamma-ray bursts. *AIP Conf. Proc.* **265**, 3–12 (eds Paciesas, W. & Fishman, G.) (AIP, New York, 1992).
- Kouveliotou, C. *et al.* Identification of two classes of gamma-ray bursts. *Astrophys. J.* **413**, L101–L104 (1993).
- Hurley, K. *et al.* Afterglow upper limits for four short-duration, hard spectrum gamma-ray bursts. *Astrophys. J.* **567**, 447–453 (2002).
- Paciesas, W. *et al.* The Fourth BATSE Gamma-Ray Burst Catalog (Revised). *Astrophys. J. Suppl.* **122**, 465–495 (1999).
- Fishman, G. *et al.* The first BATSE gamma-ray burst catalog. *Astrophys. J. Suppl.* **92**, 229–283 (1994).
- Cross, N. & Driver, S. P. The bivariate brightness function of galaxies. *Mon. Not. R. Astron. Soc.* **329**, 579–598 (2002).
- Lazzati, D., Ramirez-Ruiz, E. & Ghisellini, G. Possible detection of hard X-ray afterglows of short gamma-ray bursts. *Astron. Astrophys.* **379**, L39–L43 (2001).
- Connaughton, V. BATSE observations of gamma-ray burst tails. *Astrophys. J.* **567**, 1028–1036 (2002).
- Gebrets, N. *et al.* The Swift gamma-ray burst mission. *Astrophys. J.* **611**, 1005–1020 (2004).
- Fenimore, E. *et al.* Swift's ability to detect gamma-ray bursts. Preprint at <http://arXiv.org/astro-ph/0408513> (2004).
- Woods, P. *et al.* Evidence for a sudden magnetic field reconfiguration in soft gamma repeater SGR 1900+14. *Astrophys. J.* **552**, 748–755 (2001).
- Lyubarsky, Y., Eichler, D. & Thompson, C. Diagnosing magnetars with transient cooling. *Astrophys. J.* **580**, L69–L72 (2002).
- Lenters, G. T. *et al.* An extended burst tail from SGR 1900+14 with a thermal X-ray spectrum. *Astrophys. J.* **587**, 761–778 (2003).
- Feroci, M., Hurley, K., Duncan, R. & Thompson, C. The giant flare of 1998 August 27 from SGR 1900+14. I. An interpretive study of BeppoSAX and Ulysses observations. *Astrophys. J.* **549**, 1021–1038 (2001).
- Kulkarni, S. *et al.* The quiescent counterpart of the soft gamma repeater SGR 0526–66. *Astrophys. J.* **585**, 948–954 (2003).

51. Golenetskii, S., Ilyinskii, V. & Mazets, E. Recurrent bursts in GBS0526–66, the source of the 5 March 1979  $\gamma$ -ray burst. *Nature* **307**, 41–43 (1984).
52. Lin, R. *et al.* A three dimensional plasma and energetic particle experiment for the Wind spacecraft. *Space Sci. Rev.* **71**, 125–153 (1995).
53. Cameron, P. B. *et al.* Detection of a radio counterpart to the 27 December 2004 giant flare from SGR 1806–20. *Nature* doi:10.1038/nature03605 (this issue).

**Supplementary Information** accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We are grateful to J. Scalzo, E. Vishniac and S. Kannappan for discussions and expert help. In the US, this work was supported by NASA. The INTEGRAL mission is supported by the German government via the DLR agency.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to K.H. (khurley@ssl.berkeley.edu).



# An expanding radio nebula produced by a giant flare from the magnetar SGR 1806–20

B. M. Gaensler<sup>1</sup>, C. Kouveliotou<sup>2</sup>, J. D. Gelfand<sup>1</sup>, G. B. Taylor<sup>4,5</sup>, D. Eichler<sup>6</sup>, R. A. M. J. Wijers<sup>7</sup>, J. Granot<sup>4</sup>, E. Ramirez-Ruiz<sup>8</sup>, Y. E. Lyubarsky<sup>6</sup>, R. W. Hunstead<sup>9</sup>, D. Campbell-Wilson<sup>9</sup>, A. J. van der Horst<sup>7</sup>, M. A. McLaughlin<sup>10</sup>, R. P. Fender<sup>11</sup>, M. A. Garrett<sup>12</sup>, K. J. Newton-McGee<sup>9,13</sup>, D. M. Palmer<sup>14</sup>, N. Gehrels<sup>15</sup> & P. M. Woods<sup>3</sup>

<sup>1</sup>Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA

<sup>2</sup>NASA/Marshall Space Flight Center; and <sup>3</sup>Universities Space Research Association, NSSTC, XD-12, 320 Sparkman Drive, Huntsville, Alabama 35805, USA

<sup>4</sup>Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, PO Box 20450, Stanford, California 94309, USA

<sup>5</sup>National Radio Astronomy Observatory, PO Box O, Socorro, New Mexico 87801, USA

<sup>6</sup>Department of Physics, Ben Gurion University, POB 653, Beer Sheva 84105, Israel

<sup>7</sup>Astronomical Institute 'Anton Pannekoek', University of Amsterdam, Kruislaan 403, 1098 SJ, Amsterdam, The Netherlands

<sup>8</sup>Institute for Advanced Study, Einstein Drive, Princeton, New Jersey 08540, USA

<sup>9</sup>School of Physics, University of Sydney, New South Wales 2006, Australia

<sup>10</sup>University of Manchester, Jodrell Bank Observatory, Macclesfield, Cheshire SK11 9DL, UK

<sup>11</sup>School of Physics and Astronomy, University of Southampton, Highfield, Southampton SO17 1BJ, UK

<sup>12</sup>Joint Institute for VLBI in Europe, Postbus 2, 7990 AA Dwingeloo, The Netherlands

<sup>13</sup>Australia Telescope National Facility, CSIRO, PO Box 76, Epping, New South Wales 1710, Australia

<sup>14</sup>Los Alamos National Laboratory, PO Box 1663, Los Alamos, New Mexico 87545, USA

<sup>15</sup>NASA/Goddard Space Flight Center, Code 661, Greenbelt, Maryland 20771, USA

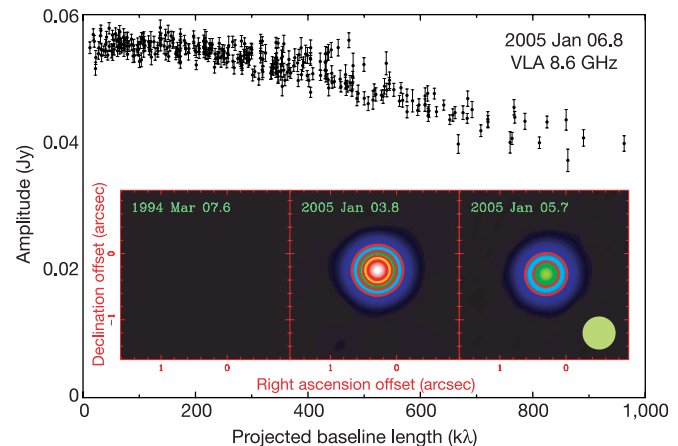
Soft  $\gamma$ -ray repeaters (SGRs) are 'magnetars', a small class of slowly spinning neutron stars with extreme surface magnetic fields,  $B \approx 10^{15}$  gauss (refs 1–3). On 27 December 2004, a giant flare<sup>4</sup> was detected from the magnetar SGR 1806–20 (ref. 2), only the third such event recorded<sup>5,6</sup>. This burst of energy was detected by a variety of instruments<sup>7,8</sup> and even caused an ionospheric disturbance in the Earth's upper atmosphere that was recorded around the globe<sup>9</sup>. Here we report the detection of a fading radio afterglow produced by this outburst, with a luminosity 500 times larger than the only other detection of a similar source<sup>10</sup>. From day 6 to day 19 after the flare from SGR 1806–20, a resolved, linearly polarized, radio nebula was seen, expanding at approximately a quarter of the speed of light. To create this nebula, at least  $4 \times 10^{43}$  ergs of energy must have been emitted by the giant flare in the form of magnetic fields and relativistic particles.

Almost seven days after the 27 December 2004 giant flare, we observed SGR 1806–20 with the Very Large Array (VLA) in its highest-resolution configuration (maximum baseline length 36.4 km). We identified a bright but fading radio source designated VLA J180839–202439 (see Fig. 1), whose position was consistent with the previously reported localization<sup>11</sup> of the SGR. This close juxtaposition, plus the transient nature of the emission, makes it certain that VLA J180839–202439 is the radio afterglow of the giant flare from SGR 1806–20. For a distance<sup>12</sup> to SGR 1806–20 of  $15d_{15}$  kiloparsecs, the 1.4-GHz flux density of this source at first detection implies an isotropic spectral luminosity of  $5d_{15}^2 \times 10^{15}$  W Hz<sup>–1</sup>, approximately 500 times larger than the radio afterglow seen from SGR 1900+14 after a giant flare in 1998 (ref. 10). No

other magnetar has been detected in the radio band, either in quiescence or during active periods<sup>13,14</sup>.

Given the very bright nature of this afterglow, we organized an international campaign over a broad range of frequencies, 0.35 to 16 GHz, to track the decay of the radio emission of VLA J180839–202439. Here we present a subset of these observations, made on days 6 to 19 after the giant flare, consisting of images made using the VLA, the Australia Telescope Compact Array (ATCA), the Westerbork Synthesis Radio Telescope (WSRT) and the Molonglo Observatory Synthesis Telescope (MOST) (see Supplementary Methods for further information).

Figure 2 shows the combined light curves from these four telescopes covering the frequency range 0.84 to 8.5 GHz. These data are consistent with a sudden increase in the decay rate at day 8.8, as summarized in Table 1 and shown by the linear fits in Fig. 2. Specifically, if we assume that  $S_\nu \propto t^\delta$  (where  $S_\nu$  is the flux density at frequency  $\nu$  and  $t$  is time in days), after day 8.8 we find an achromatic and rapid decline,  $\delta \approx -2.7$ , in six independent frequency bands (a similarly rapid decline was also observed<sup>10</sup> for the radio afterglow of SGR 1900+14 in 1998). After carefully accounting for the instrumental response of the VLA antennas we find that VLA J180839–202439 is significantly linearly polarized (see Fig. 3), which indicates that the emission mechanism is synchrotron radiation. In our earliest observations this emission was already optically thin, but showed clear evidence for a spectral steepening at

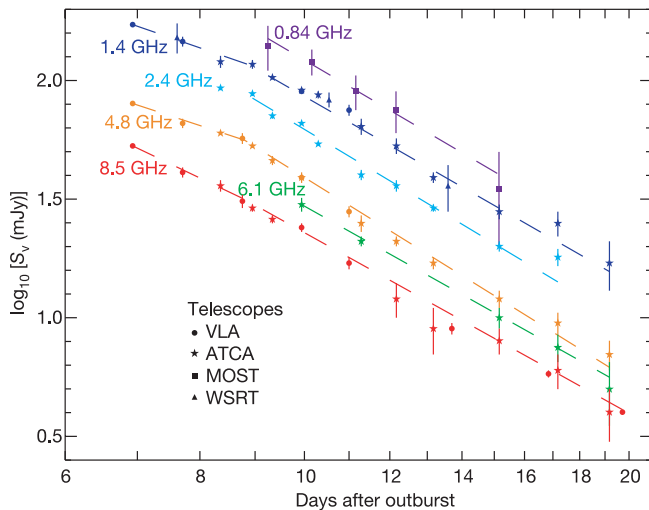


**Figure 1** Radio emission from VLA J180839–202439 at 8.5 GHz. The main panel shows the visibility amplitude as a function of projected baseline length (in units of thousands of wavelengths; 100 kλ ≈ 3.5 km) at epoch 2005 Jan 06.8 (9.9 days after the giant flare), as seen by the VLA. The data have been self-calibrated in phase until the solution converged, and each baseline has then been time-averaged over the entire observation of duration 40 min. The error bars show the standard error in the mean of the amplitude on each baseline. The decrease in amplitude as a function of increasing baseline length clearly indicates that the source is resolved. The inset shows the image of the source at three epochs, smoothed to a uniform resolution of 0.5'' (indicated by the green circle at lower right). The origin of the coordinate axes is the position of SGR 1806–20 measured with the Chandra X-ray Observatory<sup>11</sup>, which has an uncertainty of 0.3'' in each coordinate. The false-colour representation is on a linear scale, ranging from –0.3 to the peak brightness of 53 mJy beam<sup>–1</sup>. The contours are drawn at levels of 20%, 40%, 60% and 80% of this peak. No source is seen in archival 8.5-GHz data from March 1994, down to a 5σ upper limit of 0.1 mJy. In the days after the giant flare, a bright but rapidly fading source is now seen at this position. The precise location of VLA J180839–202439 was determined by phase referencing to several nearby calibrators with well-determined positions. Our best measurement was on January 16.6, for which we measured a position for VLA J180839–202439 (equinox J2000) of right ascension 18 h 08 min 39.343 ± 0.002 s, declination –20° 24' 39.80'' ± 0.04''. The source's proper motion over the time span presented in this paper is –2.8 ± 6.5 mas day<sup>–1</sup> in right ascension and –2.2 ± 6.5 mas day<sup>–1</sup> in declination.

high frequencies (see Fig. 2 legend). From day 11.2 onward, the spectrum was consistent with an unbroken power law from 0.84 to 8.5 GHz with  $\alpha = -0.75 \pm 0.02$  (where  $S_\nu \propto \nu^\alpha$ ), again similar to the 1998 afterglow of SGR 1900+14. This implies a power-law energy distribution of the emitting electrons,  $dN/dE \propto E^{-p}$ , with  $p = 1 - 2\alpha = 2.50 \pm 0.04$ .

Our highest-resolution measurements are those made with the VLA at 8.5 GHz. The visibility data from these observations, as shown for one epoch in Fig. 1, demonstrate that VLA J180839–202439 is resolved. A gaussian is a good fit to the visibilities at each epoch, with no significant persistent residuals (forthcoming higher-resolution images from the VLBA and the MERLIN array will test the validity of this model). Figure 3 demonstrates that from day 6.9 to day 19.7, the data were consistent with constant isotropic expansion since outburst at a speed  $v/c = (0.27 \pm 0.10)d_{15}$ , where  $c$  is the speed of light, with no noticeable deceleration as of day 19.7. Other than in one observation at day 16.8, the source was significantly elliptical, with an axial ratio of  $\sim 0.6$  and with the major axis oriented at about  $60^\circ$  west of north.

The spectrum and angular size of VLA J180839–202439 allow us to apply standard equipartition arguments for synchrotron sources<sup>15</sup>, implying a minimum magnetic field  $B_{\min} = 0.02d_{15}^{-2/7} [(1 + \kappa)F_{100}/f]^{2/7} \theta_{50}^{-6/7}$  gauss, where  $100F_{100}$  mJy is the



**Figure 2** Time evolution of the radio flux density from VLA J180839–202439. The x axis indicates days since the giant flare was detected from SGR 1806 20, on 2004 Dec 27.90 UT. The radio data originate from ATCA, MOST and WSRT measurements made in six independent frequency bands. Each symbol represents a different telescope, and each colour indicates a different frequency. Measurement uncertainties are indicated at the  $1\sigma$  level. Fits to the data are indicated by dashed lines, and represent the results of applying the broken power-law model described in Table 1 to the data. Significant deviations from this fit are seen at both 1.4 and 2.4 GHz, suggesting short-term time-variability in the source (most notably the possible ‘bump’ in the 1.4-GHz light curve seen with multiple telescopes on days 10–11 after the flare). These data also allow us to compute the evolution of the radio spectral index,  $\alpha$  (defined as  $S_\nu \propto \nu^\alpha$ ). At three epochs with good frequency coverage between 8.4 and 9.9 days after the flare, there is clear evidence for a spectral break, from  $\alpha \approx -0.66$  below  $\sim 5$  GHz to  $\alpha \approx -1.0$  above. Other data cannot rule out this break being present from day 6.9 (when the source was first detected) through to day 11.0. From day 11.2 onward, the spectrum has been consistent with an unbroken power law from 0.84 to 8.5 GHz with  $\alpha = -0.75 \pm 0.02$ . In addition to the data shown here, on 29 December 2004, we used the Parkes Radio Telescope at 1.4 GHz to search for radio pulsations from SGR 1806–20. For dispersion measures in the range 0 to 2,000 parsec  $\text{cm}^{-3}$ , we found no pulsed signal at or near the star’s X-ray period<sup>2</sup> of 7.5 s down to a level of  $\sim 0.2$  mJy (these data provide no constraint on unpulsed flux).

flux density of VLA J180839–202439 at 1.4 GHz,  $\kappa$  is the ratio of the energy in heavy particles to that in electrons,  $f$  is the volume-filling factor of magnetic fields and relativistic particles, and  $50\theta_{50}$  mas is the source’s angular diameter. The minimum energy in particles and magnetic fields in the emitting region is  $E_{\min} = 4 \times 10^{43} d_{15}^{17/7} [(1 + \kappa)F_{100}]^{4/7} f^{3/7} \theta_{50}^{9/7}$  erg. The spectra show no evidence of self-absorption at frequencies above 0.6 GHz at early times<sup>16</sup>, which is consistent with these parameters provided that the emitting medium has a density  $n_0 \lesssim 0.1 f \text{ cm}^{-3}$ . We can derive an additional independent energy estimate because of this rare opportunity to measure the expansion velocity directly. From the constant expansion speed observed over the first 20 days, we infer  $E_{\min} \approx 6 \times 10^{42} \Omega (n_0/0.01 \text{ cm}^{-3}) (\nu/0.27c)^5$  erg, where  $\Omega$  is the opening solid angle of the ejected material.

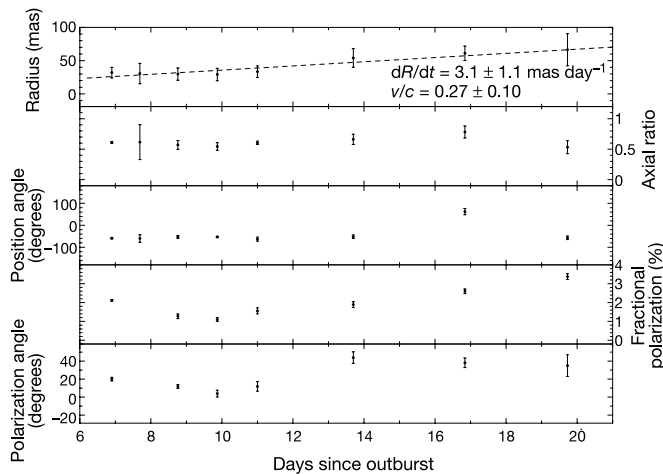
Giant flares from magnetars are thought to result from shearing and reconnection of the extreme magnetic fields near the neutron star surface<sup>17,18</sup>. The inferred minimum energy in the radio nebula is somewhat smaller than the emitted  $\gamma$ -ray energy<sup>7,8</sup>, but is much larger than the electron–positron pair luminosity that would be expected to survive annihilation close to the magnetar. This suggests that baryons may have been ablated off the surface by the intense illumination of the flaring magnetosphere<sup>17,18</sup>. The radio nebula could be naturally created by these baryons, which move off the magnetar at high velocity,  $\gtrsim 0.5c$ , and then shock the ambient medium.

The very steep decay of the radio emission after day 8.8,  $\delta \approx -2.7$ , combined with the observed subluminal expansion velocity of VLA J180839–202439, is difficult to produce in standard  $\gamma$ -ray burst blast-wave models<sup>19–21</sup>. The light curves may thus represent an adiabatically expanding population of electrons accelerated at a particularly active phase, such as might occur if the ejecta collided with a pre-existing shell. Such a shell is naturally made by SGR 1806–20 itself, because its quiescent wind<sup>22</sup> of luminosity  $\sim 10^{34} \text{ erg s}^{-1}$  will sweep up a bow shock<sup>24,25</sup> of stand-off distance  $\sim 10^{16} \text{ cm}$  (corresponding to an angular extent of  $\sim 40 d_{15}^{-1}$  mas) as it moves through the interstellar medium (ISM) at a typical neutron star velocity<sup>26</sup> of  $\sim 200 \text{ km s}^{-1}$ . The star’s motion creates a cigar-shaped cavity, mostly as a wake that trails the bow shock. If this pre-existing shell is hit by  $\sim 10^{43}$ – $10^{44}$  erg of energy from the SGR’s giant flare, it will be shocked and swept outward, resulting in a violent episode of particle acceleration that puts much of the energy into a steadily expanding synchrotron-emitting shell 4 to 8 days after the giant flare. If we suppose that this shell maintains constant thickness and constant expansion speed, then its volume,  $V$ , increases as  $t^2$ , and the magnetic field will decay as  $V^{-1/2}$  if directed within the tangent plane of the shell ( $V^{-2/3}$  if tangled in three dimensions). This predicts a power-law index for the radio decay of  $\delta = (7\alpha - 3)/3 = -2.75 \pm 0.05$  (for  $B \propto V^{-1/2}$ ) or  $\delta = (8\alpha - 4)/3 = -3.33 \pm 0.05$  (for  $B \propto V^{-2/3}$ ), consistent with the steep decay observed here. The overall evolution can be complicated by the

**Table 1** The rate of decay of the radio emission from VLA J180839–202439

| $\nu$ (GHz) | $t_0$ (days)        | $\delta_\nu (t < t_0)$ | $\delta_\nu (t > t_0)$ |
|-------------|---------------------|------------------------|------------------------|
| 0.84        | $\leq 10.2$         | ...                    | $-2.7 \pm 0.8$         |
| 1.4         | $9.0^{+0.4}_{-0.6}$ | $-1.6 \pm 0.2$         | $-2.61 \pm 0.09$       |
| 2.4         | $\leq 9.0$          | ...                    | $-2.74 \pm 0.07$       |
| 4.8         | $8.8^{+0.2}_{-0.4}$ | $-1.5 \pm 0.1$         | $-2.84 \pm 0.08$       |
| 6.1         | $\leq 11.3$         | ...                    | $-2.6 \pm 0.2$         |
| 8.5         | $8.8^{+0.2}_{-0.4}$ | $-2.2 \pm 0.2$         | $-2.54 \pm 0.04$       |

At each frequency  $\nu$ , it has been assumed that the radio flux density decays as  $S_\nu \propto t^{\delta_\nu}$ , with a break in the power-law index,  $\delta_\nu$ , at time  $t_0$ . To determine values of  $t_0$  and  $\delta_\nu$ , a weighted least-squares fit of a broken power law has been applied to each data set, with  $t_0$  a free parameter. In each case, the fit shown is the only local minimum in  $\chi^2$  that meets the requirements that there are at least two data points on either side of the break, the change in temporal index on either side of the break is larger than its uncertainties, and the power-law fits on either side of the break meet at the break point. Before day 8.8, we find that  $\delta_\nu$  possibly decreases with  $\nu$ ; after day 8.8, the flux decays rapidly at all frequencies with a power-law index  $\delta_\nu \approx -2.7$ , independent of  $\nu$ .



**Figure 3** Structural and polarization properties of VLA J180839–202439 as a function of time, as seen with the VLA at 8.5 GHz. The x axis indicates days since the giant flare. The first (top) panel plots the radius,  $R$ , of the source, determined by modelling<sup>27</sup> the visibilities at each epoch as a two-dimensional gaussian function in the Fourier plane of arbitrary position, amplitude, diameter, axial ratio and orientation, and then taking the geometric mean of the semi-major and semi-minor axes. The broken line shows a weighted linear least-squares fit to the data. The indicated expansion velocity assumes two-sided or isotropic expansion at a distance of 15 kiloparsecs. The second and third panels show the axial ratio and position angle (measured north through east) of this best-fit gaussian at each epoch. The fourth panel shows the fractional linear polarization of VLA J180839–202439 at 8.5 GHz. We find that the position angle of this linearly polarized emission is a linear function of  $\nu^{-2}$  at each epoch, indicating the presence of Faraday rotation from foreground magnetized plasma. We measure a Faraday rotation measure of  $+272 \pm 10 \text{ rad m}^{-2}$  at multiple epochs, similar to the value of  $+290 \pm 20 \text{ rad m}^{-2}$  obtained for the adjacent calibrator, MRC B1817–254, and typical of rotation measures seen through the Galactic plane<sup>28</sup>. Any contribution to the rotation measure from the immediate environment of the magnetar must thus be small. The fifth (bottom) panel shows the position angle of the electric field vector of linear polarization from VLA J180839–202439 after correction for this foreground Faraday rotation. Uncertainties at the  $1\sigma$  level are indicated for all data.

fact that the ejecta may be somewhat collimated, and may hit the shell at different places and times—we defer detailed modelling to later papers, pending higher-resolution images from MERLIN and the VLBA.

At early times, the polarization position angle on the sky is approximately perpendicular to the axis of the radio source (see Fig. 3), suggesting that the magnetic field in the emitting plasma (on average) is aligned preferentially along this axis. This is consistent with the shock producing a preferred magnetic anisotropy in the shock plane. Between observations at days 11.0 and 13.7 the polarized fraction and polarization angle both changed noticeably and a possible bump in the 1.4-GHz light curve is apparent; at day 16.8, the position angle of the major axis of the source also changed considerably. These results suggest that a different part of the outflow may have assumed the dominant role in the emission, as can occur if one region fades faster than another.

The intensity of this radio afterglow confirms the conservative inference made from the X-ray and  $\gamma$ -ray detections<sup>7,8</sup> that this event was  $\geq 2$  orders of magnitude more energetic than the 1998 giant flare from SGR 1900+14. It is difficult to attribute the difference to beaming effects, because the measured expansion velocity ( $\sim 0.3c$ ) appears to be modest. A release of  $>10^{46}$  erg in a single magnetar flare<sup>7,8</sup> suggests that a rather large fraction,  $\sim 10\%$ , of the total magnetic energy can be released at once. Continued measurements of the morphology of the expanding

radio source can provide an indication of whether the energy release took place at a specific location on the star's surface, or was a truly global phenomenon that rearranged the crust or even the entire interior. □

Received 18 January; accepted 17 February 2005; doi:10.1038/nature03498.

1. Duncan, R. C. & Thompson, C. Formation of very strongly magnetized neutron stars: Implications for gamma-ray bursts. *Astrophys. J.* **392**, L9–L13 (1992).
2. Kouveliotou, C. *et al.* An X-ray pulsar with a superstrong magnetic field in the soft  $\gamma$ -ray repeater SGR 1806–20. *Nature* **393**, 235–237 (1998).
3. Woods, P. M. & Thompson, C. in *Compact Stellar X-ray Sources* (eds Lewin, W. H. G. & van der Klis, M.) (Cambridge Univ. Press, Cambridge, in the press); preprint at (<http://arXiv.org/astro-ph/0406133>) (2004).
4. Borkowski, J. *et al.* Giant flare from SGR 1806–20 detected by INTEGRAL. *GCN Circ. No.* 2920 (2004).
5. Mazets, E. P., Golenetskii, S. V., Ilinskii, V. N., Apetkar, R. L. & Guryan, Y. A. Observations of a flaring X-ray pulsar in Dorado. *Nature* **282**, 587–589 (1979).
6. Hurley, K. *et al.* A giant periodic flare from the soft gamma repeater SGR 1900+14. *Nature* **397**, 41–43 (1999).
7. Palmer, D. M. *et al.* A giant  $\gamma$ -ray flare from the magnetar SGR 1806–20. *Nature* doi:10.1038/nature03525 (this issue).
8. Hurley, K. *et al.* An exceptionally bright flare from SGR 1806–20 and the origins of short-duration  $\gamma$ -ray bursts. *Nature* doi:10.1038/nature03519 (this issue).
9. Campbell, P. *et al.* SGR1806: Detection of a sudden ionospheric disturbance. *GCN Circ. No.* 2932 (2005).
10. Frail, D. A., Kulkarni, S. R. & Bloom, J. S. An outburst of relativistic particles from the soft gamma-ray repeater SGR 1900+14. *Nature* **398**, 127–129 (1999).
11. Kaplan, D. L. *et al.* Precise *Chandra* localization of the soft gamma-ray repeater SGR 1806–20. *Astrophys. J.* **564**, 935–940 (2002).
12. Corbel, S. & Eikenberry, S. S. The connection between W31, SGR 1806–20, and LBV 1806–20: Distance, extinction, and structure. *Astron. Astrophys.* **419**, 191–201 (2004).
13. Lorimer, D. R. & Xilouris, K. M. PSR J1907+0918: A young radio pulsar near SGR 1900+14 and G42.8+0.6. *Astrophys. J.* **545**, 385–389 (2000).
14. Kouveliotou, C. *et al.* Multiwavelength observations of the soft gamma repeater SGR 1900+14 during its 2001 April activation. *Astrophys. J.* **558**, L47–L50 (2001).
15. Pacholczyk, A. G. *Radio Astrophysics* 170–171 (Freeman, San Francisco, 1970).
16. Chandra, P. SGR 1806–20: Further low frequency GMRT results. *GCN Circ. No.* 2947 (2005).
17. Thompson, C. & Duncan, R. C. The giant flare of 1998 August 27 from SGR 1900+14. II. Radiative mechanism and physical constraints on the source. *Astrophys. J.* **561**, 980–1005 (2001).
18. Thompson, C. & Duncan, R. C. The soft gamma repeaters as very strongly magnetized neutron stars. I. Radiative mechanism for outbursts. *Mon. Not. R. Astron. Soc.* **275**, 255–300 (1995).
19. Rhoads, J. E. The dynamics and light curves of beamed gamma-ray burst afterglows. *Astrophys. J.* **525**, 737–749 (1999).
20. Granot, J., Piran, T. & Sari, R. Images and spectra from the interior of a relativistic fireball. *Astrophys. J.* **513**, 679–689 (1999).
21. Cheng, K. S. & Wang, X. Y. The radio afterglow from the giant flare of SGR 1900+14: The same mechanism as afterglows from classic gamma-ray bursts? *Astrophys. J.* **593**, L85–L88 (2003).
22. Thompson, C. *et al.* Physical mechanisms for the variable spin-down and light curve of SGR 1900+14. *Astrophys. J.* **543**, 340–350 (2000).
23. Woods, P. M. *et al.* Large torque variations in two soft gamma repeaters. *Astrophys. J.* **576**, 381–390 (2002).
24. Wilkin, F. P. Exact analytic solutions for stellar wind bow shocks. *Astrophys. J.* **459**, L31–L34 (1996).
25. Gaensler, B. M., Jones, D. H. & Stappers, B. W. An optical bow shock around the nearby millisecond pulsar J2124–3358. *Astrophys. J.* **580**, L137–L141 (2002).
26. Arzoumanian, Z., Chernoff, D. F. & Cordes, J. M. The velocity distribution of isolated radio pulsars. *Astrophys. J.* **568**, 289–301 (2002).
27. Staveley-Smith, L., Manchester, R. N., Kesteven, M. J., Tzioumis, A. K. & Reynolds, J. E. Structure in the radio remnant of Supernova 1987A. *Proc. Astron. Soc. Aust.* **10**, 331–334 (1993).
28. Brown, J. C., Taylor, A. R. & Jackel, B. J. Rotation measures of compact sources in the Canadian Galactic Plane Survey. *Astrophys. J. Suppl.* **145**, 213–223 (2003).

**Supplementary Information** accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank J. Ulvestad, J. Wrobel, R. Sault, A. Foley and R. Vermeulen for rapid scheduling of the VLA, ATCA and WSRT; T. DeLaney, G. de Bruyn and C. Brogan for assistance with data analysis; and R. Manchester, D. Frail and M. Wieringa for help with the observations. NRAO is a facility of the NSF operated under cooperative agreement by AUI. The Australia Telescope is funded by the Commonwealth of Australia for operation as a National Facility managed by CSIRO. The MOST is operated by the University of Sydney and supported in part by grants from the ARC. The WSRT is operated by ASTRON with financial support from NWO. B.M.G. acknowledges the support of NASA through a Long Term Space Astrophysics grant. D. E. acknowledges support from the Israel–US BSF, the ISF and the Arnow Chair of Physics. Y.E.L. acknowledges support from the German–Israeli Foundation. R.A.M.J.W. and A.J.H. acknowledge support from NWO.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to B.M.G. ([bgaensler@cfa.harvard.edu](mailto:bgaensler@cfa.harvard.edu)).



# A giant $\gamma$ -ray flare from the magnetar SGR 1806–20

D. M. Palmer<sup>1</sup>, S. Barthelmy<sup>2</sup>, N. Gehrels<sup>2</sup>, R. M. Kippen<sup>1</sup>, T. Cayton<sup>1</sup>, C. Kouveliotou<sup>4</sup>, D. Eichler<sup>6</sup>, R. A. M. J. Wijers<sup>7</sup>, P. M. Woods<sup>5</sup>, J. Granot<sup>8</sup>, Y. E. Lyubarsky<sup>6</sup>, E. Ramirez-Ruiz<sup>9</sup>, L. Barbier<sup>2</sup>, M. Chester<sup>10</sup>, J. Cummings<sup>2,11</sup>, E. E. Fenimore<sup>1</sup>, M. H. Finger<sup>5</sup>, B. M. Gaensler<sup>12</sup>, D. Hullinger<sup>2</sup>, H. Krimm<sup>2,3</sup>, C. B. Markwardt<sup>2,13</sup>, J. A. Nousek<sup>10</sup>, A. Parsons<sup>2</sup>, S. Patel<sup>5</sup>, T. Sakamoto<sup>2,9</sup>, G. Sato<sup>14</sup>, M. Suzuki<sup>15</sup> & J. Tueller<sup>2</sup>

<sup>1</sup>Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

<sup>2</sup>NASA/Goddard Space Flight Center, and <sup>3</sup>Universities Space Research Association, Goddard Space Flight Center, Greenbelt, Maryland 20771, USA

<sup>4</sup>NASA/Marshall Space Flight Center, and <sup>5</sup>Universities Space Research Association, NSSTC, XD-12, 320 Sparkman Dr., Huntsville, Alabama 35805, USA

<sup>6</sup>Department of Physics, Ben Gurion University, POB 653, Beer Sheva 84105, Israel

<sup>7</sup>Astronomical Institute “Anton Pannekoek”, University of Amsterdam, Kruislaan 403, 1098 SJ, Amsterdam, the Netherlands

<sup>8</sup>Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, PO Box 20450, MS 29, Stanford, California 94309, USA

<sup>9</sup>Institute for Advanced Study, Einstein Drive, Princeton, New Jersey 08540, USA

<sup>10</sup>Pennsylvania State University, University Park, Pennsylvania 16802, USA

<sup>11</sup>National Research Council, 500 Fifth St., NW Washington, DC, 20001, USA

<sup>12</sup>Harvard-Smithsonian Center for Astrophysics, 60 Garden Street MS-6, Cambridge, Massachusetts 02138, USA

<sup>13</sup>University of Maryland, College Park, Maryland 20742, USA

<sup>14</sup>Institute of Space and Astronautical Science (ISAS)/JAXA, 3-1-1 Yoshinodai, Sagami-hara, Kanagawa 229-8510, Japan

<sup>15</sup>Saitama University, 55 Shimo-Okubo, Sakura-ku, Saitama City, Saitama 338-8570, Japan

Two classes of rotating neutron stars—soft  $\gamma$ -ray repeaters (SGRs) and anomalous X-ray pulsars—are magnetars<sup>1</sup>, whose X-ray emission is powered by a very strong magnetic field ( $B \approx 10^{15}$  G). SGRs occasionally become ‘active’, producing many short X-ray bursts. Extremely rarely, an SGR emits a giant flare with a total energy about a thousand times higher than in a typical burst<sup>2–4</sup>. Here we report that SGR 1806–20 emitted a giant flare on 27 December 2004. The total (isotropic) flare energy is  $2 \times 10^{46}$  erg, which is about a hundred times higher than the other two previously observed giant flares. The energy release probably occurred during a catastrophic reconfiguration of the neutron star’s magnetic field. If the event had occurred at a larger distance, but within 40 megaparsecs, it would have resembled a short, hard  $\gamma$ -ray burst, suggesting that flares from extragalactic SGRs may form a subclass of such bursts.

Only two other giant flares have previously been recorded, one each from SGR 0526–66 on 5 March 1979 (refs 2, 3) and SGR 1900+14 on 27 August 1998 (ref. 4). Intense X-ray burst activity from SGR 1900+14 preceded the 27 August 1998 flare<sup>5</sup>; no similar activity was seen preceding the 5 March 1979 event, but it may have occurred without being detected by the instruments operating at the time, given the larger distance to SGR 0526–66 in the Large Magellanic Cloud. In the year leading up to the SGR 1806–20 flare, well-sampled X-ray monitoring observations of the source with the Rossi X-ray Timing Explorer (RXTE) indicated that it was also entering a very active phase<sup>6</sup>, emitting more frequent and intense bursts and showing enhanced persistent X-ray emission that was, indeed, a prelude to the unprecedented giant flare.

On 27 December 2004 the Swift satellite<sup>7</sup> was among a large number of spacecraft inundated by radiation from SGR 1806–20 (refs 8–10). The Burst Alert Telescope (BAT)<sup>11</sup> is a  $\gamma$ -ray (15–350 keV) coded aperture imager on Swift. Although Swift was turned away from the SGR location, and so the event illuminated the detector from behind, the flux that passed through the spacecraft and

shielding of the BAT provided excellent measurements of the event. The BAT light curve (Fig. 1; see the Supplementary Figures for more detail) demonstrates that magnetar giant flares are remarkably similar: all three start with an initial very short and spectrally hard main spike, followed by an extended softer tail highly modulated at the neutron star’s spin period. The bright, main spike lasts  $\sim 0.5$  s and is followed by a tail with  $\sim 50$  cycles of high-amplitude pulsations at the known rotation period of SGR 1806–20 (7.56 s). In the 27 December event we also notice a 1-s-long, flat-topped precursor burst at 142 s before the main spike.

Several astounding new properties of a magnetar flare are revealed from the superb time resolution of the BAT. Figure 1b plots the sharp initial rise of the main spike in time bins of 100  $\mu$ s, equivalent to the light-crossing time of the neutron star diameter. Before the steep rise of the initial spike, the count rate was rising for 40 ms at a slower rate (shown in Supplementary Fig. 2) and had reached roughly 30,000 c.p.s. (above a  $\sim 9,000$  c.p.s. background) by  $t = 0$ . At that point it increased by a factor of more than 100 in less than 1.5 ms, rising with a 0.3 ms exponential time constant. This is followed by at least one dip and continued brightening (additional dips would not be visible owing to instrument saturation) on its way to the peak. The flare rise has thus been resolved for the first time. The flux during the spike, though heavily attenuated, saturated the BAT modules, precluding a reliable flux measurement. We have therefore used the SOPA<sup>12</sup> and ESP<sup>13</sup> instruments located on geosynchronous satellites (see the Supplementary Methods) to measure the main peak flux. The SOPA instruments are small silicon detectors designed with fast event processing to measure the high particle fluxes found in orbit. During the peak of the burst, each detector had a deadtime greater than 50%, but this level of saturation can be accurately corrected for. We fitted the SOPA data with an exponential-cutoff power law (finding a characteristic temperature  $kT = 0.48(4)$  MeV and a power-law photon index  $-0.2(1)$ ) and derive a flux of  $5.0(3)$  erg cm<sup>-2</sup> s<sup>-1</sup> over an 0.160 s integration time for 45 keV to 10 MeV photons; the corresponding fluence is  $0.80(5)$  erg cm<sup>-2</sup>. (Here the number in parentheses indicates the uncertainty in the value given.) This duration and spectral hardness is in the range of characteristics found for the short, hard subclass of classic  $\gamma$ -ray bursts (GRBs)<sup>14</sup>.

Because the count rate was significantly lower during the tail, we were able to model the off-axis illumination and calibrate the flux and spectroscopy measurements (see Supplementary Methods). We find the tail of the burst to have an energy fluence of  $1.0(5) \times 10^{-3}$  erg cm<sup>-2</sup> at photon energies  $> 60$  keV. The spectral fits are consistent with thermalized spectra with  $kT \approx 15$ –30 keV as seen in previous flares, implying a comparable energy fluence below our 60-keV threshold. Accounting for the 10–60-keV photons, we project the total tail fluence to be  $\geq 2 \times 10^{-3}$  erg cm<sup>-2</sup>, roughly 0.3% that of the main peak. For a distance to SGR 1806–20 of 15  $d_{15}$  kpc (ref. 15), we then find an (isotropic equivalent) energy release of  $2 \times 10^{46} d_{15}^2$  erg in the spike, and  $5 \times 10^{43} d_{15}^2$  erg in the tail. (Here,  $d_{15} = d/(15 \text{ kpc})$ ; similarly  $E_{46} = E/(10^{46} \text{ erg})$  and  $V_{18} = V/(10^{18} \text{ cm}^2)$ .) Thus, the isotropic equivalent energy in the initial spike is about two orders of magnitude larger than that in the other two giant flares, while the energy in the tail is comparable. Indeed, a radio afterglow<sup>16</sup> was detected from this flare with a luminosity 500 times that from the 27 August flare, suggesting a very large difference in the prompt burst energy. The consistency of the tail energies among the three flares is attributable to the fact that they are limited by the storage capacity of the magnetic field<sup>1,17</sup> and should be as constant from source to source as the field energy. Thus, the tail luminosities, which are expected to be at roughly the magnetically modified Eddington limit, should also be similar, as observed. The extent of magnetic reconnection, on the other hand, governs the prompt energy release during the main spike; this

can vary greatly from one event to the next, even within the same source.

The pulse profile in the tail of the flare just after the main spike features one large peak and two smaller adjacent local maxima separated by about a quarter of a rotation cycle (Fig. 2 and Supplementary Fig. 1). The relative intensities of the peaks change during the tail, but their phases remain fixed, indicating that the field configuration does not change substantially during the tail and that the released energy comes from the trapped fireball.

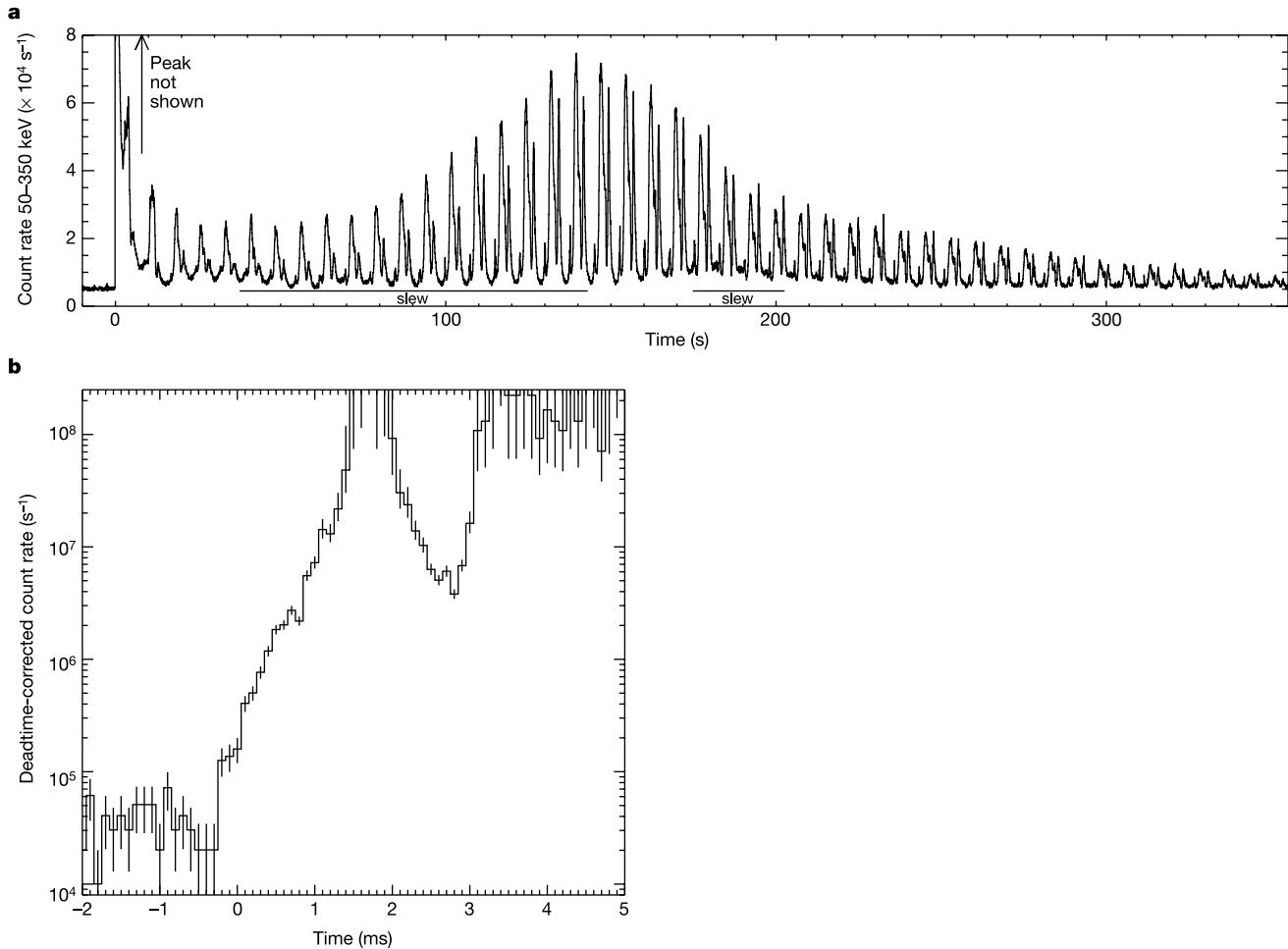
The polar  $B$  field of SGR 1806–20 has been calculated<sup>18</sup> from its spin-down rate to be  $\sim 1.6 \times 10^{15}$  G, corresponding to a external magnetic field energy of  $2 \times 10^{47}$  erg, which indicated that at most  $10d_{15}^{-2}/f$  such giant flares can be produced from the star in its lifetime (here  $f$  is the beaming factor).

We used RXTE to measure the spin frequency and spin-down rate of the SGR 30 days after the flare<sup>19</sup>. The frequency is consistent with an extrapolation of the pre-flare frequency with pre- and post-flare spin-down rates. Thus, the 27 December flare could not have caused a rapid, lasting change in the spin frequency greater than  $\sim 2 \times 10^{-5}$  Hz; this, despite the much larger apparent burst energy,

limits the frequency change to be at most comparable to that seen following the 27 August flare<sup>20</sup>. The post-flare spin-down rate,  $-3.15(9) \times 10^{-12}$  Hz s<sup>-1</sup>, although lower than it was shortly before the flare, is still in its historical range.

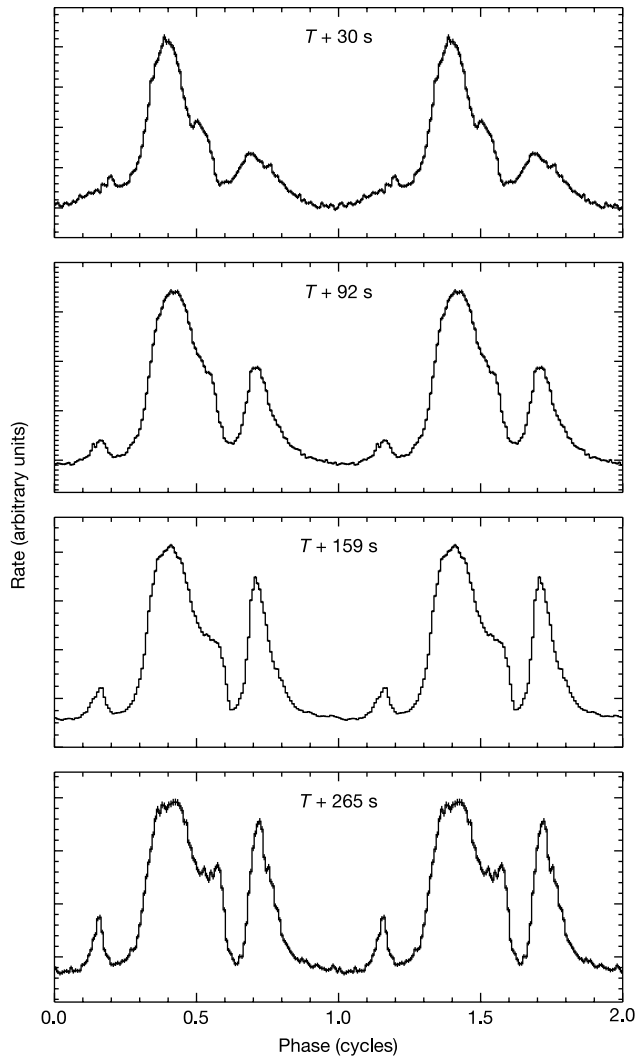
The three timescales in the phenomenon—(1) the rise time of  $\sim 1$  ms, (2) the duration of the hard spike of  $\sim 0.5$  s, and (3) the duration of the tail of several minutes—are similar for all three giant flares. These are attributed to the Alfvén propagation times in (1) the magnetosphere and (2) the star, and (3) the cooling time of the trapped pair fireball, respectively<sup>1,17</sup>.

Violent energy dissipation can occur anywhere in the magnetically dominated region, which includes the outer layers of the neutron star: if an energy of  $10^{46} E_{46}$  erg is dissipated roughly uniformly in the reconnection region of volume  $10^{18} V_{18} \text{ cm}^3$ , then matter above the layer at a density of  $10^8 E_{46}/V_{18} \text{ g cm}^{-3}$  will have an energy density larger than its gravitational potential and become unbound. This is about  $10^{24}$  g, which can be ejected into the magnetosphere at fractions of the speed of light,  $c$ . Such a mass ejection (which need not be isotropic) is enough to power the observed radio nebula and its  $0.3c$  expansion<sup>16</sup>.



**Figure 1** The SGR spike and tail light curve from BAT on Swift. **a**, BAT count rate at measured energy  $>50$  keV (64-ms bins). Although BAT was pointed  $105^\circ$  away from the SGR at the time of the main spike, it recorded  $\gamma$ -rays above 60 keV passing through and scattering within the spacecraft body and instrument shielding. As part of a pre-planned observing schedule, Swift slewed to observe a different source shortly after the main peak, reaching a steady pointing direction  $61^\circ$  from the SGR at 143 s. The spacecraft reorientation improved the detection efficiency of the SGR, visible as an apparent (not intrinsic) rise in the light curve to a peak at 140 s. This is followed by a second slew to  $67^\circ$ .

**b**, BAT deadtime-corrected count rate (all energies) during the complex leading edge of the main spike. (Note that the horizontal scale is  $10^4$  times larger than in **a**.) Uncertainties in the deadtime correction (discussed in the Supplementary Methods) make corrected count rates increasingly unreliable above  $5 \times 10^7$  counts per second. Error bars combine 1 s.d. counting statistics and the deadtime uncertainty. Time bins of 100  $\mu$ s are equivalent to the light-crossing-time of a neutron star diameter. More detailed lightcurves are shown in the Supplementary Figures.



**Figure 2** The pulse profile evolution of the magnetar SGR 1806–20 during the giant flare of 27 December 2004. Time through the flare increases from top to bottom. Each panel displays the pulse profile folded over four pulse cycles at one of four different time intervals during the flare. The times denoted at the top of each panel indicate the midpoint of the interval relative to the start of the main spike. During the first half of the tail ( $\sim 170$  s), the peak centred at phase 0.7 grows in amplitude as the primary peak fades until the two are nearly equal in height. Thereafter, the two peaks decay in lockstep while the relative amplitude of the third peak at phase 0.2 increases. Overall, the pulse profile becomes less sinusoidal during the course of the flare, that is, the power in the higher harmonics increases relative to the power at the fundamental frequency, opposite to what was seen in SGR 1900+14 during the 27 August flare<sup>24</sup>. The phases of the SGR 1806–20 pulse peaks remain fixed, indicating a finalized magnetic geometry, and supporting the notion that after the first spike no new magnetic energy is released, and only the trapped fireball energy leaks out. The individual pulses throughout the tail can be seen in Supplementary Fig. 1.

The two earlier flares would have been detectable by existing instruments only within  $\sim 8$  Mpc, and it was therefore not previously thought that such flares could be the source of the short, hard GRBs. The main spike of the 27 December giant flare would have resembled a short, hard GRB had it occurred within  $40d_{15}$  Mpc, encompassing even the Virgo cluster. Magnetar formation rate is expected to follow the star formation rate, which is (for  $z = 0$ )  $1.3 \pm 0.2 M_{\odot} \text{yr}^{-1}$  in our Galaxy<sup>21</sup> and  $0.013^{+0.02}_{-0.007} M_{\odot} \text{Mpc}^{-3} \text{yr}^{-1}$  averaged over intergalactic scales<sup>22</sup>. This suggests that the Burst And Transient Source Experiment

(BATSE) onboard the Compton Gamma Ray Observatory would have triggered on such events as short GRBs at a rate of  $N_{\text{BATSE}} = 80(\dot{N}_{\text{gal,GF}}/0.03 \text{yr}^{-1})d_{15}^3 \text{yr}^{-1}$ , to be compared to an estimate of the  $4\pi \text{sr}$  BATSE rate of about  $150 \text{yr}^{-1}$ . Here,  $\dot{N}_{\text{gal,GF}}$  is the average rate of giant flares in the Galaxy similar to the 27 December event. The observed isotropic distribution of short BATSE GRBs on the sky and the lack of excess events from the direction of the Virgo cluster suggests that only a small fraction ( $\leq 0.05$ ) of these events can be SGR giant flares within  $\leq 40$  Mpc. This implies either that  $d \leq 7$  kpc,  $N_{\text{gal,GF}} \leq 3 \times 10^{-3} \text{yr}^{-1}$  on average for a Galaxy like our own, or that the luminosity distribution includes even larger SGR flares that can be seen at a greater distance<sup>23</sup>. One possible distinction of these from the classic GRB population may well come from their radio observations, because their radio afterglows should not be detectable beyond  $\sim 1$  Mpc. The fraction of SGR events among what are now classified as short GRBs may not be predominant, but it should be detectable. This will be testable with future Swift observations.  $\square$

Received 2 February; accepted 8 March 2005; doi:10.1038/nature03525.

- Thompson, C. & Duncan, R. C. The soft gamma repeaters as very strongly magnetized neutron stars. I. Radiative mechanism for outbursts. *Mon. Not. R. Astron. Soc.* **275**, 255–300 (1995).
- Mazets, E. P., Golenetskii, S. V., Ilinskii, V. N., Apetkar, R. L. & Guryan, Y. A. Observations of a flaring X-ray pulsar in Dorado. *Nature* **282**, 587–589 (1979).
- Fenimore, E. E., Klebesadel, R. W. & Laros, J. G. The 1979 March 5 gamma-ray transient: was it a classic gamma-ray burst? *Astrophys. J.* **460**, 964–975 (1996).
- Hurley, K. et al. A giant periodic flare from the soft gamma-ray repeater SGR 1900+14. *Nature* **397**, 41–43 (1999).
- Hurley, K. et al. Reactivation and precise interplanetary network localization of the soft gamma repeater SGR 1900+14. *Astrophys. J.* **510**, L107–L109 (1999).
- Woods, P. M. et al. Gradual brightening of SGR 1806–20. *Astron. Telegr.* **313** (2004).
- Gehrels, N. et al. The SWIFT gamma-ray burst mission. *Astrophys. J.* **611**, 1005–1020 (2004).
- Borkowski, D. et al. Giant flare from SGR 1806–20 detected by INTEGRAL. *GCN Circ.* **2920** (2004).
- Palmer, D. M. et al. SGR1806–20: Swift-BAT observation of the 041227 super-flare. *GCN Circ.* **2925** (2004).
- Hurley, K. et al. An exceptionally bright flare from SGR1806–20 and the origins of short-duration  $\gamma$ -ray bursts. *Nature* doi:10.1038/nature03519 (this issue).
- Barthelmy, S. D. Burst Alert Telescope (BAT) on the Swift MIDEX mission. *SPIE* **5165**, 175–189 (2004).
- Belian, R. D., Gisler, G. R., Cayton, T. & Christensen, R. High-Z energetic particles at geosynchronous orbit during the great solar proton event series of October 1989. *J. Geophys. Res.* **97** (A11), 16897–16906 (1992).
- Meier, M. M. et al. in the *Proc. Workshop on the Earth's Trapped Particle Environment*, AIP Conf. Proc. **383**, 203–210 AIP Press, Woodbury, New York (1996).
- Kouveliotou, C. et al. Identification of two classes of gamma-ray bursts. *Astrophys. J.* **413**, L101–L104 (1993).
- Corbel, S. & Eikenberry, S. S. The connection between W31, SGR 1806–20, and LBV 1806–20: Distance, extinction, and structure. *Astron. Astrophys.* **419**, 191–201 (2004).
- Gaensler, B. M. et al. An expanding radio nebula produced by a giant flare from the magnetar SGR 1806–20. *Nature* doi:10.1038/nature03498 (this issue).
- Thompson, C. & Duncan, R. C. The giant flare of 1998 August 27 from SGR 1900+14. II. Radiative mechanism and physical constraints on the source. *Astrophys. J.* **561**, 980–1005 (2001).
- Woods, P. M. et al. Large torque variations in two soft gamma repeaters. *Astrophys. J.* **576**, 381–390 (2002).
- Woods, P. M. et al. Spin and pulsed X-ray flux properties of SGR 1806–20 after the giant flare. *Astron. Telegr.* **407** (2005).
- Woods, P. M. et al. Variable spin down in the soft gamma repeater SGR 1900+14 and correlations with burst activity. *Astrophys. J.* **524**, L55–L58 (1999).
- Kennicutt, R. C. Jr The star formation law in galactic disks. *Astrophys. J.* **344**, 685–703 (1989).
- Gallego, J., Zamorano, J., Aragon-Salamanca, A. & Rego, M. The current star formation rate of the local universe. *Astrophys. J.* **455**, L1 (1995).
- Eichler, D. Waiting for the Big One: a new class of soft gamma-ray repeater outbursts? *Mon. Not. R. Astron. Soc.* **335**, 883–886 (2002).
- Feroci, M., Hurley, K., Duncan, R. C. & Thompson, C. The giant flare of 1998 August 27 from SGR 1900+14 I. An interpretive study of *BeppoSAX* and *Ulysses* observations. *Astrophys. J.* **549**, 1021–1038 (2001).

**Supplementary Information** accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We acknowledge support from NASA (P.M.W., E.R.-R., B.M.G.); the German-Israeli Foundation (Y.E.L.); the NWO (R.A.M.J.W.); the DOE (J.G.); and the Israel-US BSF, the Israel Science Foundation, and the Arnow Chair of Theoretical Physics (D.E.).

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to D.M.P. (palmer@lanl.gov).



# Repeated injections of energy in the first 600 ms of the giant flare of SGR 1806–20

Toshio Terasawa<sup>1</sup>, Yasuyuki T. Tanaka<sup>1</sup>, Yasuhiro Takei<sup>1</sup>, Nobuyuki Kawai<sup>2</sup>, Atsumasa Yoshida<sup>3</sup>, Ken'ichi Nomoto<sup>4</sup>, Ichiro Yoshikawa<sup>1,5</sup>, Yoshifumi Saito<sup>5</sup>, Yasumasa Kasaba<sup>5</sup>, Takeshi Takashima<sup>5</sup>, Toshifumi Mukai<sup>5</sup>, Hiroto Noda<sup>6</sup>, Toshio Murakami<sup>7</sup>, Kyoko Watanabe<sup>8</sup>, Yasushi Muraki<sup>8</sup>, Takaaki Yokoyama<sup>1</sup> & Masahiro Hoshino<sup>1</sup>

<sup>1</sup>Department of Earth and Planetary Science, University of Tokyo, Tokyo 113-0033, Japan

<sup>2</sup>Department of Physics, Tokyo Institute of Technology, Tokyo 152-8551, Japan

<sup>3</sup>Department of Physics and Mathematics, Aoyama Gakuin University, Sagamihara 229-8558, Japan

<sup>4</sup>Department of Astronomy, University of Tokyo, Tokyo 113-0033, Japan

<sup>5</sup>ISAS/JAXA, Sagamihara 229-8510, Japan

<sup>6</sup>Mizusawa Astrodynamics Observatory, National Astronomical Observatory, Mizusawa 023-0861, Japan

<sup>7</sup>Department of Physics, Kanazawa University, Kanazawa 920-1192, Japan

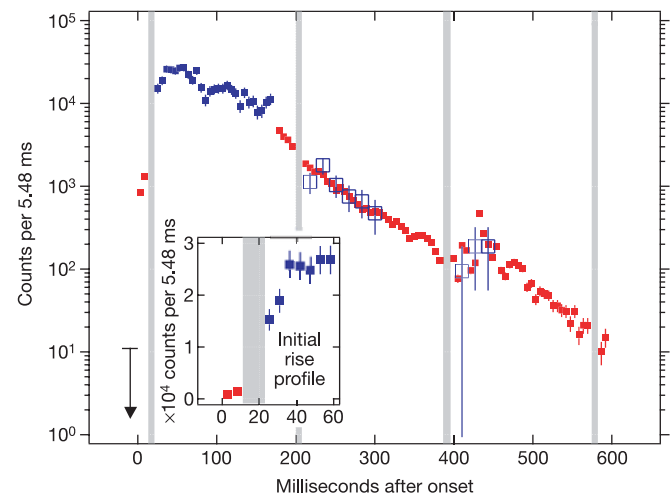
<sup>8</sup>Solar Terrestrial Environmental Research Laboratory, Nagoya University, Nagoya, Aichi 464-8601, Japan

The massive flare of 27 December 2004 from the soft  $\gamma$ -ray repeater SGR 1806–20, a possible magnetar<sup>1–3</sup>, saturated almost all  $\gamma$ -ray detectors<sup>4–7</sup>, meaning that the profile of the pulse was poorly characterized. An accurate profile is essential to determine physically what was happening at the source. Here we report the unsaturated  $\gamma$ -ray profile for the first 600 ms of the flare, with a time resolution of 5.48 ms. The peak of the profile (of the order of  $10^7$  photons  $\text{cm}^{-2} \text{s}^{-1}$ ) was reached  $\sim 50$  ms after the onset of the flare, and was then followed by a gradual decrease with superposed oscillatory modulations possibly representing repeated energy injections with  $\sim 60$ -ms intervals. The implied total energy is comparable to the stored magnetic energy in a magnetar ( $\sim 10^{47}$  erg) based on the dipole magnetic field intensity ( $\sim 10^{15}$  G), suggesting either that the energy release mechanism was extremely efficient or that the interior magnetic field is much stronger than the external dipole field<sup>2</sup>.

At the onset of the giant flare of SGR 1806–20, plasma particle detectors on the Geotail spacecraft detected an extremely strong signal of soft  $\gamma$ -ray photon fluxes (integrated above  $\sim 50$  keV) during the initial intense phase of the giant flare ( $t = 0$ –600 ms). Figure 1 shows the count profiles of two detectors,  $N_{\text{MCP}}$  (red symbols) from the microchannel plates (MCPs), and  $N_{\text{CEM}}$  (blue symbols) from channel electron multipliers (CEMs), where  $N_{\text{CEM}}$  is scaled by a factor of 280 to account for the sensitivity difference. The onset time ( $t = 0$ ) corresponded to 21 h 30 min 26.35 s Universal Time (UT), which was consistent with the expected arrival time of the onset signal at the Geotail position. To understand how the flare energy release occurred, the detailed time profile of the flare is important. Before the onset  $N_{\text{MCP}}$  was at the background level ( $< \sim 11$  counts, shown by a black arrow), and then increased to 839 counts within 5.48 ms, so that the e-fold time of the initial rise was shorter than  $\sim 1.3$  ms. After the intermediate level of 1,330 counts at  $t = 5.48$  ms, the MCPs were saturated and  $N_{\text{MCP}}$  could not be determined until  $\sim 176$  ms. Between  $t = 22.7$  and 170 ms, we could obtain  $N_{\text{CEM}}$  values instead. The scaled  $N_{\text{CEM}}$  increased to 25,900 at  $t = 33.6$  ms, thus giving an e-fold time of 9.5 ms. (A data gap between  $t = 11.0$  and 22.7 ms shown by the leftmost grey bar was due to the scheduled instrumental operation and not caused by the flare itself.) Between  $t = 33.6$  and 55.5 ms the scaled  $N_{\text{CEM}}$  stayed at the peak level of  $\sim 25,000$ – $27,000$ . After  $t = 61$  until 170 ms the scaled  $N_{\text{CEM}}$  decreased gradually with oscillatory modulation,

which suggests repeated energy injections at  $\sim 60$  ms intervals. (Note that a similar injection profile was also seen during the impulsive phase of the giant flare of SGR 0526–66 on 5 March 1979 (ref. 8).) After 176 ms  $N_{\text{MCP}}$  became available again and showed a continuing exponential decay with an e-fold time of  $\sim 66$  ms until  $t = 380$  ms. The decay profile of scaled  $N_{\text{CEM}}$  available for  $t = 210$ – $308$  ms is consistent with that of  $N_{\text{MCP}}$ . Between  $t = 397$  and 500 ms several humps were seen on the profile of  $N_{\text{MCP}}$ . Although less significant, scaled  $N_{\text{CEM}}$  showed a similar hump for  $t = 402$ – $451$  ms. (Note that the same humps were also detected by the BAT detector on the Swift spacecraft<sup>7</sup>.) The physical origin of these humps is not clear at the moment, but may represent some additional energy-releasing process. After  $t = 470$  ms,  $N_{\text{MCP}}$  again decayed with an e-fold time of  $\sim 57$  ms.

To convert the observed count rates to physical quantities such as energy flux, we need the energy spectrum information, which was not available from the Geotail observation alone. We have therefore taken three reported function forms at the peak of the giant flare<sup>4,5,7</sup> (Table 1) and integrated them above 50 keV. We then found that the resultant estimations of photon number flux, energy flux and fluence (for  $t = 0$ –600 ms) are almost independent of the choice of the energy spectrum, and are  $\sim 2.5 \times 10^7$  photons  $\text{cm}^{-2} \text{s}^{-1}$ ,  $\sim 20$  erg  $\text{s}^{-1} \text{cm}^{-2}$  and  $\sim 2$  erg  $\text{cm}^{-2}$ , respectively. The corresponding total energy radiated from SGR 1806–20 is estimated to be  $\sim 5 \times 10^{46} (\Omega/4\pi) d_{15}^2$  erg, where  $d_{15}$  is the distance scaled by 15 kpc and  $\Omega$  is the solid angle of the radiation. Here we note that the solid-angle factor ( $\Omega/4\pi$ ) is not likely to be as small as  $10^{-2}$ , as is typically assumed for relativistic jets for GRBs. ( $\Omega/4\pi > 0.1$  is more likely for the intense initial spikes of SGRs because they have been seen in all the three giant flares of SGRs that could have been detected without them. Therefore the presence of a very efficient mechanism is implied, which promptly releases (on a timescale of  $\sim 60$  ms) a major fraction of the stored magnetic energy in a magnetar,  $E_{\text{mag}} \approx 1.7 \times 10^{47} B_{15}^2 R_6^3$  erg (where  $B_{15}$  and  $R_6$  are the internal magnetic field scaled by  $10^{15}$  G and the radius of the magnetar scaled by  $10^6$  cm). Alternatively, as suggested by ref. 2,

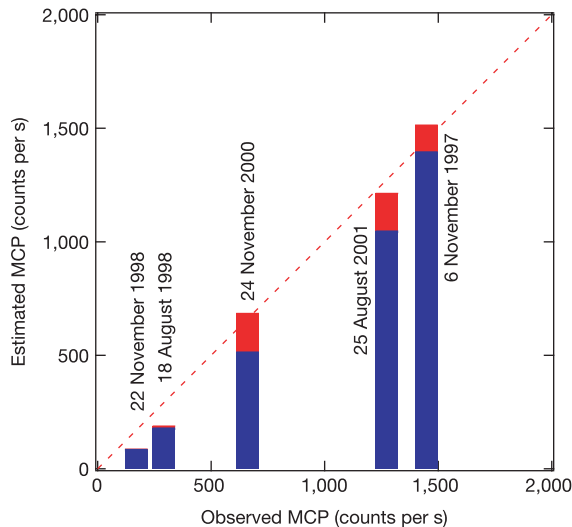


**Figure 1** Observed photon counts during the first 600 ms of the giant flare.  $N_{\text{MCP}}$  (red squares) and  $N_{\text{CEM}}$  (blue solid squares) show the counts of MCP and CEM instruments accumulated over a bin of 5.48-ms duration. Averages over three successive bins (16.44 ms) are taken for  $N_{\text{CEM}}$  (blue open squares) after 200 ms. Vertical error bars represent  $1\sigma$  statistical deviations, which become smaller than the symbol size for  $N_{\text{MCP}} > 1,000$ . Grey bars show the data gaps every 32 bins (or every 187.09 ms) caused by the scheduled instrumental operation. Before  $t = 0$ , the photon counts were below  $\sim 11$ . The black arrow shows this upper limit. The inset shows  $t = 0$ –60 ms with a linear scale.

Table 1 Estimations of energy flux, fluence, luminosity and total energy for several models

| Model                   | Photon flux<br>(photons cm <sup>-2</sup> s <sup>-1</sup> ) | Energy flux<br>(erg cm <sup>-2</sup> s <sup>-1</sup> ) | Fluence<br>(erg cm <sup>-2</sup> ) | Isotropic luminosity<br>(erg s <sup>-1</sup> ) | Isotropic total energy<br>(erg)               |
|-------------------------|------------------------------------------------------------|--------------------------------------------------------|------------------------------------|------------------------------------------------|-----------------------------------------------|
| Planck                  | $(2.5^{+1.1}_{-0.6}) \times 10^7$                          | $19^{+9}_{-4}$                                         | $2.0^{+0.9}_{-0.5}$                | $(5.1^{+2.3}_{-1.2}) \times 10^{47} d_{15}^2$  | $(5.4^{+2.4}_{-1.3}) \times 10^{46} d_{15}^2$ |
| $E^{-0.2} \exp(-E/480)$ | $(2.5^{+0.9}_{-0.6}) \times 10^7$                          | $18^{+7}_{-4}$                                         | $1.9^{+0.7}_{-0.45}$               | $(4.9^{+1.7}_{-1.0}) \times 10^{47} d_{15}^2$  | $(5.2^{+1.8}_{-1.1}) \times 10^{46} d_{15}^2$ |
| $E^{-0.7} \exp(-E/800)$ | $(2.5^{+0.8}_{-0.5}) \times 10^7$                          | $18^{+6}_{-3}$                                         | $1.9^{+0.6}_{-0.4}$                | $(4.9^{+1.4}_{-1.0}) \times 10^{47} d_{15}^2$  | $(5.1^{+1.6}_{-0.9}) \times 10^{46} d_{15}^2$ |

The Planck distribution with  $kT = 175$  keV is from ref. 4; the power-law distributions with exponential cut-offs at 480 and 800 keV are from refs 7 and 5, respectively.



**Figure 2** Calibration of the  $\gamma$ -ray sensitivity of the MCP detector. The observed MCP counts (counts per s) are compared with the counts synthesized from the solar  $\gamma$ -ray photon observations with the estimated  $[eS]_{MCP,L}$  and  $[eS]_{MCP,H}$  for five solar  $\gamma$ -ray flares (Goes class X3.7 on 22 November 1998, X4.9 on 18 August 1998, X2.3 on 24 November 2000, X5.3 on 25 August 2001 and X9.4 on 6 November 1997). The blue and red parts of each bar represent contributions from the photons in the energy ranges of 50–500 keV and above 500 keV, respectively.

the internal magnetic field could be as strong as  $(5\text{--}10) \times 10^{15}$  G so as to permit the emission of multiple giant flares over the lifetime of a magnetar.

As we noted above, there were humps in the light curve of the 2004 giant flare at  $t = 400\text{--}500$  ms. Similar humps were also observed<sup>9</sup> at 200–600 ms after the onset of another giant flare of SGR 1900+14 on 27 August 1998, whose total energy is smaller by a factor of  $\sim 100$  than the 2004 giant flare. From similarities in the timings of the humps despite the large difference in the total energies, we suggest that the humps more probably represented continuing energy injections, rather than the results of interactions of the flare ejecta with environmental matter.  $\square$

## Methods

The Low Energy Particle (LEP) experiment<sup>10</sup> aboard Geotail consists of an ion detector with seven independent MCPs and electronics systems, and an electron detector with seven independent CEMs and electronics systems, both of which are designed to measure plasma particles in the solar wind and magnetospheric environment. When the giant flare occurred, Geotail was at  $(-1.5997 \times 10^5, -97,945, -19,671)$  km using the Geocentric Solar Inertial (GCI) coordinates (J2000), which was in the solar wind about  $\sim 10$  earth radii upstream from the bow shock. Although MCPs and CEMs kept measuring the solar wind ions and electrons throughout the giant flare interval, these particles were being selected electrostatically and came into the detectors mainly at some limited timings that fortunately did not overlap the giant flare interval: the contribution of solar wind ions to  $N_{MCP}$ , which is the sum of counts over seven MCPs, was at most 30, and thus was negligible for the study of intense  $\gamma$ -ray photons. On the other hand, during subintervals ( $t < 10$  ms,  $175$  ms  $< t < 200$  ms,  $320$  ms  $< t < 400$  ms and  $450$  ms  $< t < 600$  ms),  $N_{CEM}$ , which is the sum of counts over seven CEMs, was affected by solar wind electrons, and was not available for the  $\gamma$ -ray photon detection.

Another fortunate factor was that the angular distance between the Sun and

SGR 1806–20 was  $\sim 5$  degrees, so that previous knowledge of the ‘calibration’ of the LEP detector to be used as a soft  $\gamma$ -ray photon counter on the basis of solar flare photon analysis was directly applicable to the interpretation of the observed characteristics of photons from SGR 1806–20. By comparing the count rates of MCP and CEM with the hard X-ray<sup>11</sup> and  $\gamma$ -ray<sup>12</sup> data from the Yohkoh space solar observatory during major solar flares in 1997–2001 (ref. 13), we have seen that MCP and CEM are sensitive to soft  $\gamma$ -ray photons above  $\sim 50$  keV, where their sensitivities are evaluated as the product  $[eS]$  of quantum efficiency  $\varepsilon$  and effective detection area  $S$  summed over seven MCPs and seven CEMs against  $\gamma$ -ray photons. (Here  $\varepsilon$  is defined to include not only the detector response itself but also the attenuation factor inside the spacecraft.) From the spectral information provided by the Yohkoh observations, we have calculated photon fluxes in two energy ranges, L (50–500 keV) and H (above 500 keV), and then estimated  $[eS]_{MCP,L}$  and  $[eS]_{MCP,H}$  separately for these two energy ranges. (This separation is possible because the energy spectra of incident solar  $\gamma$ -ray photons differ from event to event.) Figure 2 shows the calibration result. Along the line of sight towards SGR 1806–20, we have found that  $[eS]_{MCP,L} = (0.19 \pm 0.06)$  cm<sup>2</sup> and  $[eS]_{MCP,H} = (0.22 \pm 0.16)$  cm<sup>2</sup> where systematic errors are included. The estimation of  $[eS]_{CEM}$  has been also done with solar flare  $\gamma$ -ray photons, and the result is summarized as  $[eS]_{CEM} \approx 1/280$  of  $[eS]_{MCP}$ . The smallness of  $[eS]_{CEM}$  as compared with  $[eS]_{MCP}$  is consistent with the difference in physical sizes of CEMs (millimetres) and MCPs (several centimetres).

It is noted that  $[eS]_{MCP}$  obtained above is by a factor  $10^{-2}\text{--}10^{-3}$  smaller than those of conventional  $\gamma$ -ray detectors. Nonetheless MCPs were saturated during  $\sim 150$  ms of the onset of the giant flare of SGR 1806–20. We have made that the dead-time analysis for MCPs and found that the characteristic dead time is  $\sim 4.3$   $\mu$ s, which is consistent with the pre-flight calibration of the LEP system as well as the calculated circuit time constant.  $N_{MCP}$  shown in Fig. 1, is after the dead-time correction, which becomes significant above  $\sim 1,000$  counts. On the other hand, CEMs, which are two orders of magnitude less sensitive than MCPs, were found to be free from the saturation effect even at the peak of the giant flare.

Received 14 February; accepted 23 March 2005; doi:10.1038/nature03573.

- Thompson, C. & Duncan, R. C. The soft gamma repeaters as very strongly magnetized neutron stars – I. Radiative mechanism for outbursts. *Mon. Not. R. Astron. Soc.* **275**, 255–300 (1995).
- Thompson, C. & Duncan, R. C. The soft gamma repeaters as very strongly magnetized neutron stars. II. Quiescent neutrino, X-ray, and Alfvén wave emission. *Astrophys. J.* **473**, 322–342 (1996).
- Thompson, C. & Duncan, R. C. The giant flare of 1998 August 27 from SGR1900+14. II. Radiative mechanism and physical constraints on the source. *Astrophys. J.* **561**, 980–1005 (2001).
- Hurley, K. et al. An exceptionally bright flare from SGR1806–20 and the origins of short-duration  $\gamma$ -ray bursts. *Nature* doi:10.1038/nature03519 (this issue); preprint at <http://arXiv.org/astro-ph/0502329> (2005).
- Mazets, E. P. et al. The Konus-Wind and Helicon-Coronas-F detection of the giant  $\gamma$ -ray flare from the soft  $\gamma$ -ray repeater SGR 1806–20. Preprint at <http://arXiv.org/astro-ph/0502541> (2005).
- Mereghetti, S. et al. The first giant flare SGR 1806–20: observations with the INTEGRAL SPI anti-coincidence shield. *Astrophys. J. Lett.* (in the press); preprint at <http://arXiv.org/astro-ph/0502577> (2005).
- Palmer, D. M. et al. A giant  $\gamma$ -ray flare from the magnetar SGR 1806–20. *Nature* doi:10.1038/nature03525 (this issue); preprint at <http://arXiv.org/astro-ph/0503030> (2005).
- Barat, C. et al. Fine time structure in the 1979 March 5 gamma ray burst. *Astron. Astrophys.* **126**, 400–402 (1983).
- Mazets, E. P. et al. Activity of the soft gamma repeater SGR 1900+14 in 1998 from Konus-Wind observations: 2. The giant August 27 outburst. *Astron. Lett.* **25**, 635–648 (1999).
- Mukai, T. et al. The low energy particle (LEP) experiment onboard the Geotail satellite. *J. Geomag. Geoelectr.* **46**, 669–692 (1994).
- Kosugi, T. et al. The hard X-ray telescope (HXT) for the solar-A mission. *Solar Phys.* **136**, 17–36 (1991).
- Yoshimori, M. et al. The wide band spectrometer on the solar-A. *Solar Phys.* **136**, 69–88 (1991).
- Matsumoto, Y. et al. A statistical study of gamma-ray emitting solar flares observed with Yohkoh. *Publ. Astron. Soc. Jpn.* **57**, 211–220 (2005).

**Acknowledgements** We thank R. Yamazaki, K. Ioka, K. Shibata, T. Minoshima and K. Hurley for comments and discussions. This work was supported in part by grants from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to T.T. (terasawa@eps.s.u.tokyo.ac.jp).

# Detection of a radio counterpart to the 27 December 2004 giant flare from SGR 1806–20

P. B. Cameron<sup>1</sup>, P. Chandra<sup>3,4</sup>, A. Ray<sup>3</sup>, S. R. Kulkarni<sup>1</sup>, D. A. Frail<sup>5</sup>, M. H. Wieringa<sup>6</sup>, E. Nakar<sup>2</sup>, E. S. Phinney<sup>2</sup>, Atsushi Miyazaki<sup>7</sup>, Masato Tsuboi<sup>8</sup>, Sachiko Okumura<sup>8</sup>, N. Kawai<sup>9</sup>, K. M. Menten<sup>10</sup> & F. Bertoldi<sup>11</sup>

<sup>1</sup>Division of Physics, Mathematics and Astronomy, 105-24; and <sup>2</sup>Theoretical Astrophysics, 130-33, California Institute of Technology, Pasadena, California 91125, USA

<sup>3</sup>Tata Institute of Fundamental Research, Mumbai 400 005, India

<sup>4</sup>Joint Astronomy Programme, Indian Institute of Science, Bangalore 560 012, India

<sup>5</sup>National Radio Astronomy Observatory, Socorro, New Mexico 87801, USA

<sup>6</sup>Australia Telescope National Facility, CSIRO, PO Box 76, Epping, New South Wales 1710, Australia

<sup>7</sup>Shanghai Astronomical Observatory 80, Nandan Road, Shanghai 200030, China

<sup>8</sup>Nobeyama Radio Observatory, National Astronomical Observatory of Japan, Minamiasaku, Nagano 384-1305, Japan

<sup>9</sup>Department of Physics, Tokyo Institute of Technology, Ookayama 2-12-1, Meguro-ku, Tokyo 152-8551, Japan

<sup>10</sup>Max Planck Institut für Radioastronomie, Auf dem Hügel 69; and

<sup>11</sup>University of Bonn, Auf dem Hügel 71, 53121 Bonn, Germany

It was established over a decade ago that the remarkable high-energy transients known as soft  $\gamma$ -ray repeaters (SGRs) are located in our Galaxy<sup>1,2</sup> and originate from neutron stars with intense ( $\leq 10^{15}$  G) magnetic fields—so-called ‘magnetars’<sup>3</sup>. On 27 December 2004, a giant flare<sup>4</sup> with a fluence<sup>5</sup> exceeding  $0.3 \text{ erg cm}^{-2}$  was detected from SGR 1806–20. Here we report the detection of a fading radio counterpart to this event. We began a monitoring programme from 0.2 to 250 GHz and obtained a high-resolution 21-cm radio spectrum that traces the intervening interstellar neutral hydrogen clouds. Analysis of the spectrum yields the first direct distance measurement of SGR 1806–20: the source is located at a distance greater than 6.4 kpc and we argue that it is nearer than 9.8 kpc. If correct, our distance estimate lowers the total energy of the explosion and relaxes the demands on theoretical models. The energetics and the rapid decay of the radio source are not compatible with the afterglow model that is usually invoked for  $\gamma$ -ray bursts. Instead, we suggest that the rapidly decaying radio emission arises from the debris ejected during the explosion.

On 3 January 2005 we observed SGR 1806–20 with the Very Large Array (VLA) and identified and promptly reported<sup>6</sup> a new radio source at right ascension  $\alpha_{J2000} = 18 \text{ h } 08 \text{ min } 39.34 \text{ s}$  and declination  $\delta_{J2000} = -20^\circ 24' 39.7''$  (with an uncertainty of  $\pm 0.1''$  in each coordinate) coincident with the quiescent X-ray counterpart<sup>7</sup>. In Table 1 we report the results of a subsequent monitoring programme undertaken with the VLA, the Giant Metre-wave Radio Telescope (GMRT), the Australia Telescope Compact Array (ATCA), the Nobeyama Millimeter Array (NMA) and the Institut de Radioastronomie Millimétrique (IRAM) 30-m telescope.

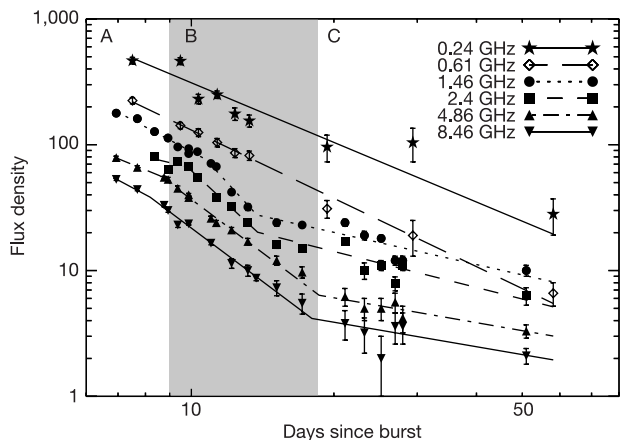
The radio source decays in all frequency bands, but its behaviour is complex (Fig. 1). At each band we model the flux with a power law,  $S_\nu(t) \propto t^\alpha$ , but allow for changes in the temporal indices  $\alpha$  (‘breaks’) at two epochs. These breaks are clearly seen in our highest signal-to-noise ratio data. After the first break (nine days postburst) the light curve steepens to  $-4 \leq \alpha \leq -3$ . The radio source<sup>8</sup> from SGR 1900+14 following the 27 August 1998 giant flare<sup>9</sup> showed a similar rapid decay at 8 GHz. Subsequently, around day 14, the light curve flattens to  $\alpha \approx -1$ . At any given epoch, the radio spectrum can be modelled by a power law,  $S_\nu \propto \nu^\beta$ . The spectral index,  $\beta$ ,

steepens with time, changing from about  $-0.7$  to  $-0.9$  (see Fig. 1 and Supplementary Information).

We confirm claims that the source is resolved<sup>10</sup> by an independent analysis. We find that it is elongated with a major-axis of  $\theta \approx 77$  milliarcseconds (mas) and an axial ratio of 2:1 (Table 2). We considered four expansion models:  $\theta \propto t^s$  with unconstrained  $s$ , and three plausible models ( $s = 0, 2/5$  and 1). The best-fit model corresponds to no expansion ( $s = 0.04 \pm 0.15$ ). However, owing to the limited range of our observations we prefer not to model the dynamics of the explosion.

We took advantage of the brightness of the radio source and obtained a high-resolution spectrum (Fig. 2b) centred around the 21-cm line of atomic hydrogen (H I). Intervening interstellar clouds appear as absorption features in the spectrum. These clouds are expected to participate in the rotation of the Galaxy and the absorption features allow us to infer ‘kinematic’ distance estimates. Such estimates have several caveats. First, in the inner Galaxy the radial velocity curve is double-valued (see Fig. 2c) leading to a ‘near’ distance estimate ( $d_i$ ) and a ‘far’ distance estimate ( $d_u$ ) for each velocity. Second, in some directions, there are features with non-circular motion, for example, the ‘3-kpc expanding arm’ and the ‘ $-30 \text{ km s}^{-1}$  spiral arm’<sup>11</sup>. Finally, in the innermost part of the Galaxy there is a deficit of cold gas<sup>12</sup>.

Significant H I absorption towards SGR 1806–20 is seen over the velocity range  $-20$  to  $+85 \text{ km s}^{-1}$  (Fig. 2b). There is also a weak ( $2.5\sigma$ ) absorption feature coincident in velocity with a clearly detected  $^{12}\text{CO}(1-0)$  emission feature identified<sup>13</sup> as MC94 (Fig. 2a). Adopting a simple galactic rotation curve with a circular velocity



**Figure 1** Broadband temporal behaviour of the transient radio source coincident with SGR 1806–20. The abscissa indicates days elapsed since the giant flare on 27.90 December 2004. The displayed flux density measurements (denoted with symbols) were obtained in six frequency bands with the VLA, GMRT and ATCA (Table 1). The error bars denote  $1\sigma$  uncertainties. With the exception of the 6.1-GHz data (which is insufficiently sampled at early and late times and is not shown), the light curves with  $\nu > 1 \text{ GHz}$  are best fitted by power-law models (shown as lines,  $S_\nu \propto t^{\alpha_\nu}$ ) with two breaks at  $t_1 \sim 9$  days and  $t_2 \sim 15$  days (see the Supplementary Information for exact values). The temporal index varies chromatically in the time before and after the first break (denoted by regions A and B respectively). The exponent value ranges from  $-2 \leq \alpha_A \leq -1$  and  $-4 \leq \alpha_B \leq -3$ ; here the subscript identifies the region of interest. After day  $\sim 15$  (region C) the source decay flattens to  $\alpha_C \approx -0.9$  at these frequencies, which persists until day 51. Region B, the period of steep light-curve decline, is shaded grey. The light curves with  $\nu < 1 \text{ GHz}$  do not show these temporal breaks or late time flattening. Apparently a single power-law decay model with  $\alpha = -1.57 \pm 0.11$  (0.24 GHz) and  $\alpha = -1.80 \pm 0.08$  (0.61 GHz) provides a good statistical description of the data. Our substantial frequency coverage (over three decades) allows an excellent characterization of the spectrum. The spectrum is consistent with a single power-law slope ( $S_\nu \propto \nu^\beta$ ) at all epochs. On day 7, before the first temporal break, we find  $\beta = -0.62 \pm 0.02$ . The spectrum steepens to a value of  $\beta = -0.76 \pm 0.05$  (day 15), reaching  $\beta = -0.9 \pm 0.1$  (days 21–51).



Table 1 Flux densities of SGR 1806–20

| Epoch          | Telescope | $\Delta t^*$<br>(days) | $S_{0.24}$<br>(mJy) | $S_{0.610}$<br>(mJy) | $S_{1.46}$<br>(mJy) | $S_{2.4}$<br>(mJy) | $S_{4.86}$<br>(mJy) | $S_{6.1}$<br>(mJy) | $S_{8.46}^\dagger$<br>(mJy) | $S_{102}$<br>(mJy) |
|----------------|-----------|------------------------|---------------------|----------------------|---------------------|--------------------|---------------------|--------------------|-----------------------------|--------------------|
| 3.84 Jan 2005  | VLA       | 6.94                   | –                   | –                    | 178 ± 4             | –                  | 79 ± 2              | –                  | 53 ± 1                      | –                  |
| 4.17 Jan 2005  | NMA       | 7.27                   | –                   | –                    | –                   | –                  | –                   | –                  | –                           | 16.3 ± 5.6         |
| 4.41 Jan 2005  | GMRT      | 7.51                   | 466 ± 28            | 224 ± 13             | –                   | –                  | –                   | –                  | –                           | –                  |
| 4.59 Jan 2005  | VLA       | 7.69                   | –                   | –                    | 161 ± 4             | –                  | 66 ± 2              | –                  | 44 ± 1                      | –                  |
| 5.26 Jan 2005  | ATCA      | 8.36                   | –                   | –                    | 127 ± 3             | 80 ± 2             | –                   | –                  | –                           | –                  |
| 5.66 Jan 2005  | VLA       | 8.76                   | –                   | –                    | –                   | –                  | 55 ± 1              | –                  | 33 ± 1                      | –                  |
| 5.85 Jan 2005  | ATCA      | 8.93                   | –                   | –                    | 113 ± 3             | 63 ± 2             | 53 ± 2              | –                  | 30 ± 1                      | –                  |
| 6.26 Jan 2005  | ATCA      | 9.36                   | –                   | –                    | 96 ± 3              | 73 ± 2             | 45 ± 2              | –                  | 23 ± 1                      | –                  |
| 6.38 Jan 2005  | GMRT      | 9.48                   | 462 ± 29            | 142 ± 8              | –                   | –                  | –                   | –                  | –                           | –                  |
| 6.77 Jan 2005  | VLA       | 9.87                   | –                   | –                    | 93 ± 2              | –                  | 38 ± 1              | –                  | 23.5 ± 0.5                  | –                  |
| 6.77 Jan 2005  | ATCA      | 9.87                   | –                   | –                    | 85 ± 3              | 67 ± 2             | 40 ± 1              | 32 ± 1             | –                           | –                  |
| 7.20 Jan 2005  | ATCA      | 10.30                  | –                   | –                    | 88 ± 2              | 55 ± 1             | –                   | –                  | –                           | –                  |
| 7.25 Jan 2005  | GMRT      | 10.35                  | 231 ± 20            | 125 ± 9              | –                   | –                  | –                   | –                  | –                           | –                  |
| 7.90 Jan 2005  | VLA       | 11.00                  | –                   | –                    | 71 ± 2              | –                  | 26 ± 1              | –                  | 16.5 ± 0.5                  | –                  |
| 8.19 Jan 2005  | ATCA      | 11.29                  | –                   | –                    | 67 ± 3              | 38 ± 2             | 24 ± 1              | 20 ± 1             | –                           | –                  |
| 8.24 Jan 2005  | GMRT      | 11.34                  | 250 ± 17            | 104 ± 8              | –                   | –                  | –                   | –                  | –                           | –                  |
| 9.06 Jan 2005  | ATCA      | 12.16                  | –                   | –                    | 42 ± 2              | 32 ± 1.5           | 21 ± 1              | –                  | 11.4 ± 1                    | –                  |
| 9.26 Jan 2005  | GMRT      | 12.36                  | 176 ± 20            | 86 ± 7               | –                   | –                  | –                   | –                  | –                           | –                  |
| 10.07 Jan 2005 | ATCA      | 13.16                  | –                   | –                    | 32 ± 2              | 24 ± 1             | 17 ± 1              | –                  | 10 ± 1                      | –                  |
| 10.16 Jan 2005 | GMRT      | 13.26                  | 155 ± 17            | 82 ± 7               | –                   | –                  | –                   | –                  | –                           | –                  |
| 10.60 Jan 2005 | VLA       | 13.70                  | –                   | –                    | –                   | –                  | –                   | –                  | 8.7 ± 0.4                   | –                  |
| 12.00 Jan 2005 | NMA       | 15.10                  | –                   | –                    | –                   | –                  | –                   | –                  | –                           | 7.16 $\dagger$     |
| 12.04 Jan 2005 | ATCA      | 15.14                  | –                   | –                    | 24 ± 1.5            | 16 ± 1             | 12 ± 1              | 9.3 ± 1            | 7.3 ± 1                     | –                  |
| 13.00 Jan 2005 | NMA       | 16.10                  | –                   | –                    | –                   | –                  | –                   | –                  | –                           | 5.50 $\dagger$     |
| 14.04 Jan 2005 | ATCA      | 17.14                  | –                   | –                    | 23 ± 1              | 15 ± 1             | 9.7 ± 1             | 7.3 ± 1            | 5.5 ± 1                     | –                  |
| 16.25 Jan 2005 | GMRT      | 19.35                  | 96 ± 23             | 31 ± 5               | –                   | –                  | –                   | –                  | –                           | –                  |
| 16.37 Jan 2005 | GMRT      | 19.47                  | –                   | –                    | 20 ± 2 $\S$         | –                  | –                   | –                  | –                           | –                  |
| 18.01 Jan 2005 | ATCA      | 21.11                  | –                   | –                    | 24 ± 1.5            | 17 ± 1             | 6.2 ± 1             | 4.7 ± 1            | 3.8 ± 1                     | –                  |
| 20.10 Jan 2005 | ATCA      | 23.20                  | –                   | –                    | 19 ± 1.5            | 10 ± 1.5           | 5 ± 1               | –                  | 3.2 ± 1                     | –                  |
| 22.07 Jan 2005 | ATCA      | 25.17                  | –                   | –                    | 18 ± 1              | 11 ± 1             | 5 ± 1               | 4.3 ± 1            | 2.0 ± 1                     | –                  |
| 23.84 Jan 2005 | ATCA      | 26.94                  | –                   | –                    | 12 ± 1              | 7.9 ± 1            | 5.6 ± 1             | 3.7 ± 1            | 3.6 ± 1                     | –                  |
| 24.85 Jan 2005 | ATCA      | 27.95                  | –                   | –                    | 12 ± 1              | 11 ± 1             | 4.2 ± 1             | 5.1 ± 1            | 3.6 ± 1                     | –                  |
| 26.26 Jan 2005 | GMRT      | 29.36                  | 104 ± 31            | 19 ± 6               | –                   | –                  | –                   | –                  | –                           | –                  |
| 4.01 Feb 2005  | GMRT      | 38.14                  | –                   | –                    | 10 $\dagger$ , $\S$ | –                  | –                   | –                  | –                           | –                  |
| 16.87 Feb 2005 | ATCA      | 50.97                  | –                   | –                    | 10 ± 1              | 6.3 ± 1            | 3.3 ± 0.4           | –                  | 2.1 ± 0.3                   | –                  |
| 24.01 Feb 2005 | GMRT      | 58.11                  | 28.2 ± 9            | 6.6 ± 1.4            | –                   | –                  | –                   | –                  | –                           | –                  |

Flux density measurements of the transient radio counterpart to SGR 1806–20 from the VLA, GMRT, NMA and ATCA as a function of frequency and time. The reported errors are  $1\sigma$ . In addition to these measurements, we obtained IRAM-30 m observations on 8 and 9 January 2005 using MAMBO-2 at 250 GHz, which show no detection with a value of  $0.57 \pm 0.46$  mJy at the position of the radio source. Finally, we detect linearly polarized emission from the source at the 1.5% to 2.5% level. See the Supplementary Information for observational details.

\*The epoch of the flare,  $t_0$ , was 27.90 December 2004.

$\dagger$ ATCA observations in this column have a frequency of 8.6 GHz.

$\S$ These values represent  $2\sigma$  upper limits.

$\S$ The frequency is 1.06 GHz for the 16.37 January 2005 and 4.01 February 2005 GMRT observations.

$\Theta_0 = 220 \text{ km s}^{-1}$  and a Galactic Centre distance  $R_0 = 8.5 \text{ kpc}$ , the near distance to SGR 1806–20 (for  $V_{\text{LSR}} = 95 \text{ km s}^{-1}$ ) is  $d_1 = 6.4 \text{ kpc}$ .

The two H I emission clouds seen at velocities above  $100 \text{ km s}^{-1}$  towards SGR 1806–20 (Fig. 2a), with no corresponding H I absorption, may be used to infer an upper limit to the distance, provided that we can be reasonably certain that cold neutral gas exists at these velocities. The H I absorption spectrum towards the nearby ( $\Delta\theta = 0.77^\circ$ ) extragalactic source J1811–2055 shows a strong and broad absorption feature between  $110$  and  $130 \text{ km s}^{-1}$  (Fig. 2d). The only H I emission in this direction<sup>14</sup> above  $60 \text{ km s}^{-1}$  corresponds to an H I cloud at the same velocity. This feature can be traced in absorption towards several other extragalactic radio sources in this direction<sup>15</sup> suggesting that cold gas at  $\sim 120 \text{ km s}^{-1}$

is widespread. Adopting the same galactic rotation curve as above, the absorbing cloud at  $+120 \text{ km s}^{-1}$  can either be located at 7 or 9.8 kpc (see Fig. 2c). We thus suggest an upper limit to the distance,  $d_u = 9.8 \text{ kpc}$ .

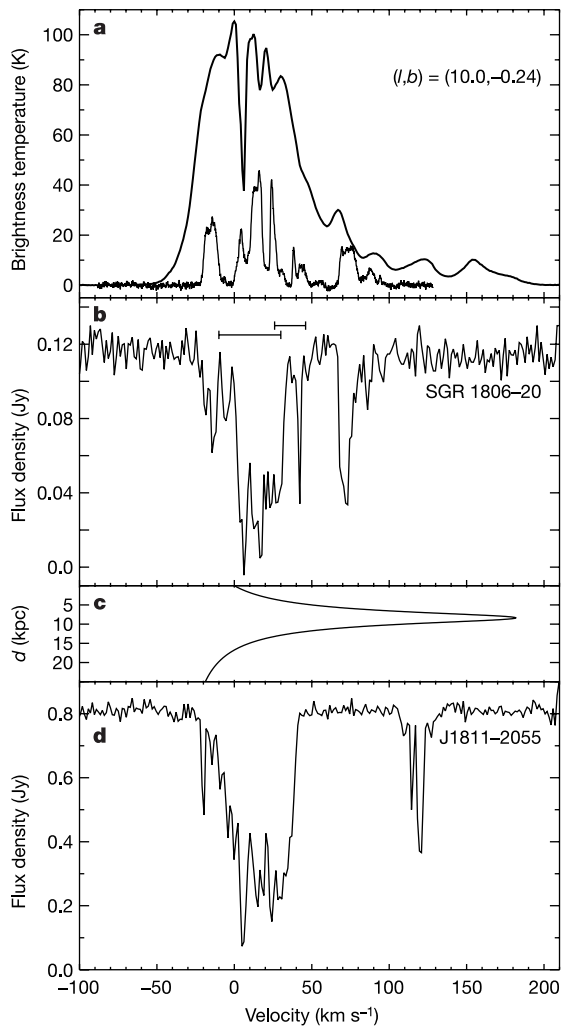
Our new distance estimate is smaller than previous (indirect) values<sup>11,16</sup> of 12 to 15 kpc. Accepting our estimate has several important implications. It results in a reduction of the total energy released ( $\propto d^2$ ) as well as the rate of such events in nearby galaxies<sup>17</sup> ( $\propto d^3$ ), and calls into question the association<sup>18</sup> of SGR 1806–20 with a star cluster along the same line of sight. Therefore, claims that magnetars originate from more massive stars than normal neutron stars<sup>19</sup> may be called into question.

Next we consider the energetics of the material giving rise to the radio emission. As with many other radio sources, the power-law

Table 2 Size measurements of the radio source

| Epoch       | Beam<br>(mas) | Beam position angle<br>(degrees) | Fit major axis<br>(mas) | Axial ratio             | Fit position angle<br>(degrees) |
|-------------|---------------|----------------------------------|-------------------------|-------------------------|---------------------------------|
| 03 Jan 2005 | 349 × 170     | 11.8                             | 78.2 $^{+3.0}_{-2.9}$   | 0.34 $^{+0.21}_{-0.34}$ | 54.6 $^{+6.7}_{-6.0}$           |
| 04 Jan 2005 | 593 × 173     | –40.3                            | 72.4 $^{+14.5}_{-48}$   | 0.00 $^{+0.90}_{-0}$    | 69 $^{+20}_{-67}$               |
| 05 Jan 2005 | 397 × 168     | –25.7                            | 55 $^{+18}_{-10}$       | 0.66 $^{+0.34}_{-0.66}$ | 74 $^{+90}_{-90}$               |
| 06 Jan 2005 | 329 × 178     | –16.5                            | 75.7 $^{+3.0}_{-3.0}$   | 0.48 $^{+0.18}_{-0.33}$ | 51.8 $^{+9.0}_{-8.8}$           |
| 07 Jan 2005 | 532 × 178     | 40.4                             | 78 $^{+26}_{-18}$       | 0.60 $^{+0.4}_{-0.60}$  | 69 $^{+90}_{-90}$               |
| 10 Jan 2005 | 560 × 161     | –38.8                            | 112 $^{+30}_{-42}$      | 0.34 $^{+0.33}_{-0.34}$ | 18 $^{+28}_{-18}$               |

Sizes and 95% confidence limits of the radio source as measured with the VLA at 8.46 GHz. The source is clearly resolved at all VLA epochs. The best constraints come from the observations that occurred closest to the transit of the source on 3 and 6 January. The result is a source with size  $\theta \approx 77 \text{ mas}$ , an axial ratio of  $\sim 0.5$  and a position angle of 60 degrees (measured clockwise from the North). The flux centroid did not change position within the limits of our astrometric accuracy ( $\pm 100 \text{ mas}$ ). The best-fit model is consistent with no expansion,  $s = 0.04 \pm 0.15$  ( $\theta(t) \propto t^s$ ) with a  $\chi^2 = 5.3$  with four degrees of freedom. The Sedov–Taylor ( $s = 2/5$ ) and free-expansion ( $s = 1$ ) models were also fitted, and yield  $\chi^2 = 20.6$  and  $\chi^2 = 91$ , respectively. These fits have five degrees of freedom. See the Supplementary Information for the details of the source size measurements.



**Figure 2** Cold atomic and molecular hydrogen spectra towards SGR 1806–20. These spectra were used to derive a distance estimate for SGR 1806–20. **a**, H I emission (upper curve) in the direction of SGR 1806–20, determined by averaging two adjacent spectra taken by ref. 14 at  $l, b = (10.0^\circ, 0.0^\circ)$  and  $l, b = (10.0^\circ, -0.5^\circ)$ . The lower curve is the  $^{12}\text{CO}(1-0)$  spectrum (from ref. 11). For display purposes, the brightness temperature has been scaled up by a factor of 11.4 for the lower curve. **b**, The H I absorption spectrum taken towards SGR 1806–20. The two horizontal bars illustrate the radial velocity measurements<sup>16,11</sup> of the nearby star LBV 1806–20 ( $36 \pm 10$  and  $10 \pm 20 \text{ km s}^{-1}$ ). The absorption spectra were made with the Very Large Array on 4 January 2005, using a 1.56-MHz bandwidth in both hands of polarization, centred at  $50 \text{ km s}^{-1}$  with respect to the local standard of rest. The bandwidth was divided into 256 channels each 6.1 kHz in width, or a velocity resolution of  $1.3 \text{ km s}^{-1}$  covering a velocity range of  $-115$  to  $+215 \text{ km s}^{-1}$ . **c**, The distance as a function of radial velocity adopting a simple Galactic rotation curve with  $\Theta_0 = 220 \text{ km s}^{-1}$  and  $R_0 = 8.5 \text{ kpc}$ . **d**, The H I absorption spectrum of nearby extragalactic source J1811–2055 at  $l, b = (9.8^\circ, -1.0^\circ)$ . The lower limit to the distance is firmly established by the  $95 \text{ km s}^{-1}$  absorption feature from MC94 (see text). An upper limit to the distance of the SGR is suggested by the absence of strong absorption at  $+120 \text{ km s}^{-1}$ , seen towards J1811–2055 and several other extragalactic radio sources in this direction<sup>15</sup>. One could argue that the  $120 \text{ km s}^{-1}$  cold cloud is small and not present along the line of sight to SGR 1806–20. However, this hypothesis also requires the absence of any other cloud between  $95 \text{ km s}^{-1}$  (distance of 6.4 kpc) and  $86 \text{ km s}^{-1}$  (distance of 10.6 kpc). The mean absorption coefficient drops in the inner Galaxy ( $R \lesssim 4.5 \text{ kpc}$ ), giving a mean free path between clouds of 2.3 kpc (ref. 15). The distance interval from 6.4 to 10.6 kpc corresponds to  $\sim 1.8$  mean free paths. So, the probability of finding no clouds in this gap is 16%. Thus, our upper limit of 9.8 kpc is not a certainty, but is quite likely.

spectrum can be attributed to energetic electrons with a power-law energy distribution ( $dN/d\gamma \propto \gamma^{-p}$ ; here  $\gamma$  is the Lorentz factor of electrons and we measure  $p = 2.24 \pm 0.04$  on day 7, which is a typical value for strong shocks) that gyrate in a magnetic field and emit synchrotron radiation. We apply the minimum energy formulation for synchrotron sources<sup>20,21</sup> to the radio spectrum (from 0.2 to 100 GHz) of 3 January 2005 and find the energy of the radio-emitting source and the associated magnetic field strength are  $U_{\min} \approx 10^{43} d_{10}^{17/7} \theta_{75}^{9/7} \text{ erg}$  and  $B_{\min} \approx 13 d_{10}^{-2/7} \theta_{75}^{-6/7} \text{ mG}$ ; here, the distance is  $10 d_{10} \text{ kpc}$  and the angular diameter is  $75 \theta_{75} \text{ mas}$  (Table 2).

Evidently, the amount of energy released in the  $\gamma$ -ray flare,  $E_{\gamma, \text{iso}} \geq 4 \times 10^{45} \text{ erg s}^{-1}$  (assuming unbeamed, isotropic emission), substantially exceeds  $U_{\min}$ . In contrast, the ratio  $U_{\min}/E_{\gamma, \text{iso}}$  is unity for  $\gamma$ -ray bursts and as a result the lower-energy and longer-duration emission is correctly regarded as arising from the shock of the circumburst medium (the ‘afterglow’ model). Thus, based solely on energetics, there is at first sight no reason to suggest that the radio source is the afterglow of the  $\gamma$ -ray flare.

Furthermore, as discussed above, the radio emission decays quite rapidly nine days after the burst. Such a rapid decay is incompatible with the afterglow model (in the non-relativistic limit), for which we expect<sup>22</sup>  $\alpha = 3\beta + 0.6$ . We conclude (in contrast to refs 23, 24 and 25) that the radio emission must be powered by something very different from that which produced the  $\gamma$ -ray emission.

In summary, the radio emission can be described by two components: a rapidly decaying component and a slowly decaying component. The latter becomes detectable when the former has decayed significantly. The rapid decay is phenomenologically similar to that seen from accreting Galactic sources (for example, ref. 26)—the so-called ‘plasmon’ model framework, in which the radio emission arises from a ball of electrons and magnetic field which are initially shocked and then cool down by expansion. We make the specific suggestion that the radio emission, up until about two weeks, is a result of the shocking of the debris given off in the explosion (the ‘reverse shock’). In this framework, the slowly decaying component is the emission arising from the forward shock as the ejecta slams into the circumburst medium. A requirement of this suggestion is that the energy inferred in the slowly decaying component should be comparable to  $U_{\min}$ . Separately, we note that the comparable ratios  $U_{\min}/E_{\gamma, \text{iso}}$  and the temporal and spectral similarities of the giant flares from SGR 1806–20 and SGR 1900+14 suggest a common mechanism for launching these flares and similar circumstellar environments.

Regardless of the suggestions and speculations, it is clear that the radio afterglow is telling us something entirely different from that revealed by the  $\gamma$ -ray emission. If our suggestion of a reverse-shock origin is correct, then radio observations allow us to probe the ejecta. Taken together, it appears that rapid and intense radio monitoring of such flares will be highly fruitful in the future.  $\square$

Received 9 February; accepted 5 April 2005; doi:10.1038/nature03605.

- Kulkarni, S. R. & Frail, D. A. Identification of a supernova remnant coincident with the soft gamma-ray repeater SGR1806–20. *Nature* **365**, 33–35 (1993).
- Murakami, T. et al. X-Ray identification of the soft gamma-ray repeater 1806–20. *Nature* **368**, 127–128 (1994).
- Woods, P. M. & Thompson, C. Soft gamma repeaters and anomalous x-ray pulsars: Magnetar candidates. In *Compact Stellar X-ray Sources* (eds Lewin, W. H. G. & van der Klis, M.) (Cambridge Astrophysics Series, Cambridge Univ. Press, Cambridge, in the press); preprint at <http://arXiv.org/astro-ph/0406133> (2004).
- Borkowski, J. et al. Giant flare from SGR 1806–20 detected by INTEGRAL. *GRB Circ. Network* **2920** (2004).
- Boggs, S. et al. SGR 1806–20, RHESSI observations of the 041227 giant flare. *GRB Circ. Network* **2936** (2005).
- Cameron, P. B. & Kulkarni, S. R. VLA observations of SGR 1806–20. *GRB Circ. Network* **2928** (2005).
- Kaplan, D. L. et al. Precise Chandra localization of the soft gamma-ray repeater SGR 1806–20. *Astrophys. J.* **564**, 935–940 (2002).
- Frail, D. A., Kulkarni, S. R. & Bloom, J. S. An outburst of relativistic particles from the soft gamma-ray repeater SGR 1900+14. *Nature* **398**, 127–129 (1999).
- Feroi, M., Hurley, K., Duncan, R. C. & Thompson, C. The giant flare of 1998 August 27 from SGR

- 1900+14. I. An interpretive study of BeppoSAX and Ulysses observations. *Astrophys. J.* **549**, 1021–1038 (2001).
10. Gaensler, B. M. *et al.* Second-epoch VLA observations of SGR 1806–20. *GRB Circ. Network* **2933** (2005).
11. Corbel, S. & Eikenberry, S. S. The connection between W31, SGR 1806–20, and LBV 1806–20: Distance, extinction, and structure. *Astron. Astrophys.* **419**, 191–201 (2004).
12. Kolpak, M. A., Jackson, J. M., Bania, T. M. & Dickey, J. M. The radial distribution of cold atomic hydrogen in the galaxy. *Astrophys. J.* **578**, 868–876 (2002).
13. Corbel, S. *et al.* The distance of the soft gamma repeater SGR 1806–20. *Astrophys. J.* **478**, 624–630 (1997).
14. Hartmann, D. & Burton, W. B. *Atlas of Galactic Neutral Hydrogen*. Ch. 4, 169 (Cambridge Univ. Press, Cambridge, 1997).
15. Garwood, R. W. & Dickey, J. M. Cold atomic gas in the inner Galaxy. *Astrophys. J.* **338**, 841–861 (1989).
16. Figer, D. E., Najarro, F. & Kudritzki, R. P. The double-lined spectrum of LBV 1806–20. *Astrophys. J.* **610**, L109–L112 (2004).
17. Nakar, E., Gal-Yam, A., Piran, T., Fox, D. B. The distances of short-hard GRBs and the SGR connection. *Astrophys. J.* (submitted); preprint at <http://arXiv.org/astro-ph/0502148> (2005).
18. Fuchs, Y. *et al.* ISO observations of the environment of the soft gamma-ray repeater SGR 1806–20. *Astron. Astrophys.* **350**, 891–899 (1999).
19. Gaensler, B. M. *et al.* A stellar wind bubble coincident with the anomalous X-ray pulsar 1E 1048.1–5937: Are magnetars formed from massive progenitors? *Astrophys. J.* **620**, L95–L98 (2005).
20. Pacholczyk, A. G. *Radio Astrophysics. Nonthermal Processes in Galactic and Extragalactic Sources* (Series of Books in Astronomy and Astrophysics, Freeman, San Francisco, 1970).
21. Scott, M. A. & Readhead, A. C. S. The low-frequency structure of powerful radio sources and limits to departures from equipartition. *Mon. Not. R. Astron. Soc.* **180**, 539–550 (1977).
22. Frail, D. A., Waxman, E. & Kulkarni, S. R. A 450-day light curve of the radio afterglow of GRB 970508: Fireball calorimetry. *Astrophys. J.* **537**, 191–204 (2000).
23. Cheng, K. S. & Wang, X. Y. The radio afterglow from the giant flare of SGR 1900+14: The same mechanisms as afterglows from classic gamma-ray bursts? *Astrophys. J.* **593**, L85–L88 (2003).
24. Nakar, E., Piran, T., Sari, R. Giant flares as mini gamma ray bursts. Preprint at <http://arXiv.org/astro-ph/0502052> (2005).
25. Wang, X. Y., Wu, X. F., Fan, Y. Z., Dai, Z. G. & Zhang, B. An energetic blast wave from the December 27 giant flare of soft  $\gamma$ -ray repeater 1806–20. *Astrophys. J.* **623**, L29–L32 (2005).
26. Hjellming, R. M. *et al.* Light curves and radio structure of the 1999 September transient event in V4641 Sagittarii (= XTE J1819–254 = SAX J1819.3–2525). *Astrophys. J.* **544**, 977–992 (2000).

Supplementary Information accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** ATCA is funded by the Commonwealth of Australia for operations as a National Facility managed by CSIRO. We thank K. Newton-McGee and B. Gaensler for scheduling and performing observations with the ATCA. GMRT is run by the National Centre for Radio Astrophysics—Tata Institute of Fundamental Research, India. We thank the GMRT staff and in particular C. H. Ishwara-Chandra and D. V. Lal for help with observations and analysis. The VLA is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc. NMA is a branch of the National Astronomical Observatory, National Institutes of Natural Sciences, Japan. IRAM is supported by INSU/CNRS (France), MPG (Germany) and IGN (Spain). We thank A. Weiss from IRAM for help with the observations. We gratefully acknowledge discussions with S. Corbel, S. S. Eikenberry and R. Sari. Our work is supported in part by the NSF and NASA.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to P.B.C. ([pbcc@astro.caltech.edu](mailto:pbcc@astro.caltech.edu)).

## Observation of nuclear fusion driven by a pyroelectric crystal

B. Naranjo<sup>1</sup>, J.K. Gimzewski<sup>2,3</sup> & S. Puterman<sup>1,3</sup>

<sup>1</sup>Physics Department, <sup>2</sup>Chemistry Department, <sup>3</sup>CNSI, University of California Los Angeles, California 90095, USA

While progress in fusion research continues with magnetic<sup>1</sup> and inertial<sup>2</sup> confinement, alternative approaches—such as Coulomb explosions of deuterium clusters<sup>3</sup> and ultrafast laser–plasma interactions<sup>4</sup>—also provide insight into basic processes and technological applications. However, attempts to produce fusion in a room temperature solid-state setting, including ‘cold’ fusion<sup>5</sup> and ‘bubble’ fusion<sup>6</sup>, have met with deep scepticism<sup>7</sup>. Here we report that gently heating a pyroelectric crystal in a deuterated atmosphere can generate fusion under desktop conditions. The

electrostatic field of the crystal is used to generate and accelerate a deuteron beam (>100 keV and >4 nA), which, upon striking a deuterated target, produces a neutron flux over 400 times the background level. The presence of neutrons from the reaction  $D + D \rightarrow {}^3\text{He}$  (820 keV) + n (2.45 MeV) within the target is confirmed by pulse shape analysis and proton recoil spectroscopy. As further evidence for this fusion reaction, we use a novel time-of-flight technique to demonstrate the delayed coincidence between the outgoing  $\alpha$ -particle and the neutron. Although the reported fusion is not useful in the power-producing sense, we anticipate that the system will find application as a simple palm-sized neutron generator.

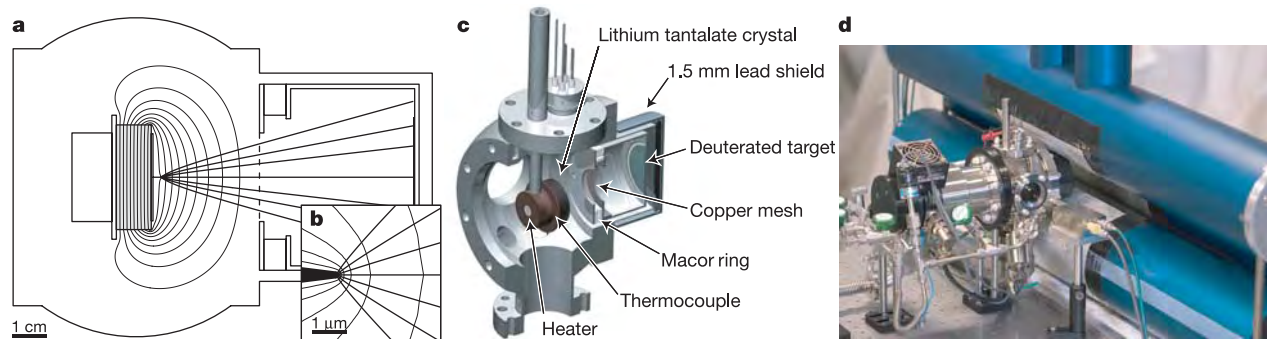
Because its spontaneous polarization is a function of temperature, heating or cooling a pyroelectric crystal in vacuum causes bound charge to accumulate on faces normal to the polarization. A modest change in temperature can lead to a surprisingly large electrostatic field. For example, heating a lithium tantalate crystal from 240 K to 265 K decreases its spontaneous polarization by  $0.0037 \text{ C m}^{-2}$  (ref. 8). In the absence of spurious discharges, introducing this magnitude of surface charge density into the particular geometry of our experiment (Fig. 1a, b) gives a potential of 100 kV. Attempts to harness this potential have focused on electron acceleration and the accompanying bremsstrahlung radiation<sup>9–12</sup>, but using the crystal to produce and accelerate ions has been studied much less. Seeking to drive the D–D fusion reaction (<http://www.physics.ucla.edu/~naranjo/ucei/ucei.pdf>; <http://neer.inel.gov/abstract.asp?ProjectID=126>; <http://www.binghamton.edu/physics/Brownridge%20Summary.pdf>), we set out to develop a method of reliably producing an ion beam of sufficient energy (>80 keV) and current (>1 nA). We demonstrate such a method using a tungsten tip to generate the high field (>25 V nm<sup>−1</sup>) necessary for gas phase field ionization of deuterium.

A cut-away view of our vacuum chamber is shown in Fig. 1c. We mounted a cylindrical (diameter, 3.0 cm; height, 1.0 cm) z-cut LiTaO<sub>3</sub> crystal with negative axis facing outward onto a hollow copper block. On the exposed crystal face, we attached a copper disc (diameter, 2.5 cm; height, 0.5 mm), allowing charge to flow to a tungsten probe (shank diameter, 80  $\mu\text{m}$ ; tip radius, 100 nm; length, 2.3 mm) (Fig. 1b). The probe geometry was chosen so that the tip field was approximately  $25 \text{ V nm}^{-1}$  when the crystal face was charged to 80 kV.

Our detector arrangement is shown in Fig. 1d. The neutron detector consists of six liquid scintillator (BC-501A and NE213) cells (diameter, 127 mm; height, 137 mm), each optically coupled to a 127-mm Hamamatsu R1250 photomultiplier tube (PMT). One output of each PMT was fed into a logical OR trigger, while the other output was fed into two Acqiris DC270 8-bit (1 gigasample per second) 4-channel digitizers configured as a single 8-channel digitizer. For every trigger, a 650-ns waveform was digitized simultaneously on all channels and written to disk for later analysis.

A typical run is shown in Fig. 2. The chamber’s deuterium pressure was held at 0.7 Pa throughout the run. First, the crystal was cooled down to 240 K from room temperature by pouring liquid nitrogen into the cryogenic feedthrough. At time  $t = 15 \text{ s}$ , the heater was turned on. At  $t = 100 \text{ s}$ , X-ray hits due to free electrons striking the crystal were recorded. At  $t = 150 \text{ s}$ , the crystal had reached 80 kV and field ionization was rapidly turning on. At  $t = 160 \text{ s}$  and still not above  $0^\circ\text{C}$ , the neutron signal rose above background. Ions striking the mesh and the surrounding aperture created secondary electrons that accelerated back into the crystal, increasing the X-ray signal. At  $t = 170 \text{ s}$ , the exponential growth of the ion current had ceased, and the tip was operating in the strong field regime, in which neutral molecules approaching the tip ionize with unity probability. The neutron flux continued to increase along with crystal potential until  $t = 220 \text{ s}$ , when we shut off the heater. Then, the crystal lost charge through field ionization faster than the





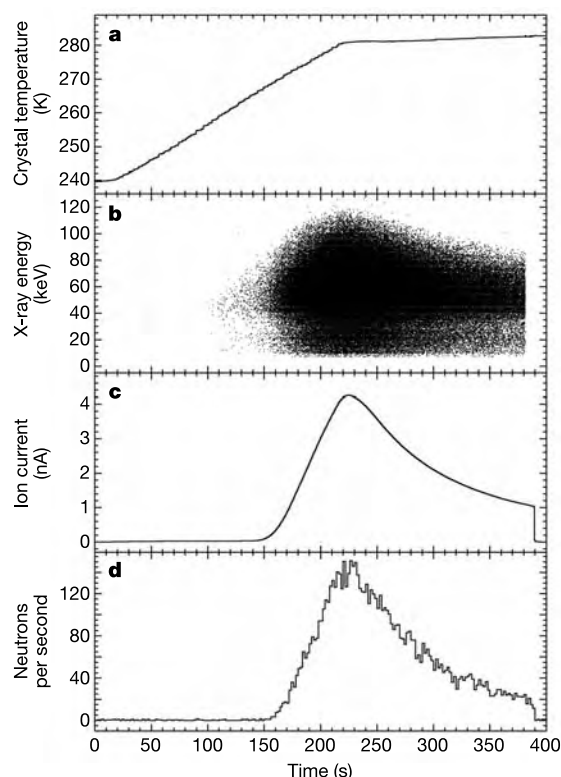
**Figure 1** Experiment geometry. **a**, Calculated equipotentials and D<sup>+</sup> trajectories for a crystal charged to 100 kV; calculations were performed using finite-element methods. The grounded copper mesh (85% open area, 19.8-μm wire; vertical dashed line) shields the Faraday cup (right). The cup and target are connected to a Keithley 6485 picoammeter and biased to +40 V to collect secondary electrons and help prevent avalanche discharges. **b**, Same trajectories shown near the tip. Using a shorter tip reduces the beam's angular spread. **c**, Vacuum chamber cut-away view. D<sub>2</sub> pressure was set using a

leak valve and monitored with a D<sub>2</sub> compensated Pirani gauge. The target was a molybdenum disc coated with ErD<sub>2</sub>. **d**, Arrangement of neutron and X-ray detectors (Amptek XR-100T-CdTe). To better resolve the bremsstrahlung endpoint, a 2.5-cm aluminium filter (not shown) was placed between the X-ray detector and the viewport. The vacuum chamber's thick stainless steel walls and lead sheet shielded the neutron detector from X-rays.

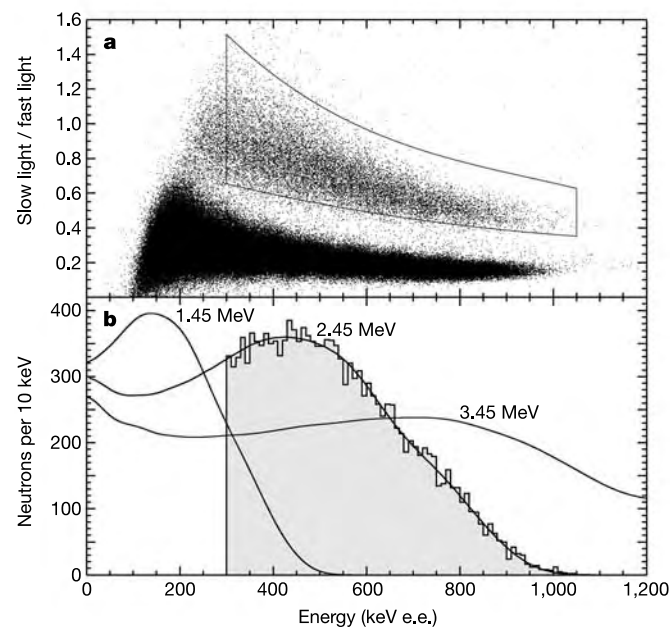
reduced pyroelectric current could replace it, resulting in a steadily decreasing crystal potential. At  $t = 393$  s, the crystal spontaneously discharged by sparking, halting the effect.

Pulse shape analysis and proton recoil spectroscopy of neutron detector data collected during the run are shown in Fig. 3. (See Supplementary Methods for details on neutron detector calibration, pulse shape analysis, and a Monte Carlo calculation of detector response and efficiency.) The majority of background triggers, as collected in the first 100 s of the run, have an electron recoil shape

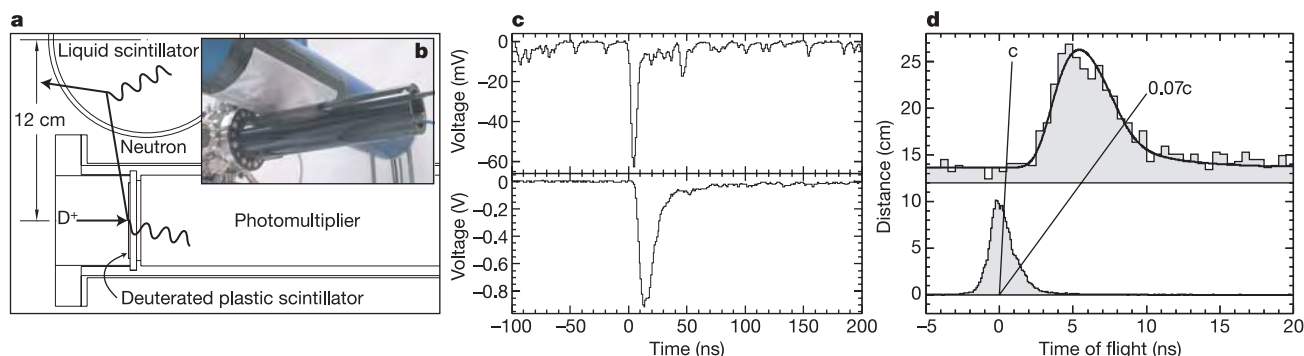
(900 counts per second) and are due to cosmic muons and  $\gamma$ -rays, compared with relatively few triggers having a proton recoil shape (33 counts in the first 100 s). Correcting for our 18% 2.45-MeV neutron detection efficiency, the observed peak neutron flux was 800 neutrons per second. We may compare this observed peak



**Figure 2** Data from a single run (see also Supplementary Movie 1). **a**, Crystal temperature. The heating rate was 12.4 K min<sup>-1</sup>, corresponding to a pyroelectric current of 22 nA and a heating power of 2 W. **b**, X-rays detected. **c**, Faraday cup current. **d**, Neutrons detected.



**Figure 3** Neutron spectroscopy for the single run. The energy scale, given in electron equivalent (e.e.) energy, was calibrated against Compton edges of a series of  $\gamma$ -ray sources and is proportional to anode charge. **a**, Pulse shape discrimination (PSD) spectrum. Our PSD variable 'slow light/fast light' is the ratio of integrated light in the tail of the PMT signal generated by an event in the liquid scintillator, to the integrated light around the signal's peak. Electron recoils are in the lower branch, and proton recoils, having longer scintillation decay, are in the upper branch. The events enclosed within the upper region are compared against tabulated pulse shapes, rejecting unusual events such as PMT double pulsing. There were a total of 15,300 valid neutrons over the course of the 400-s run. From the distribution of events, we estimate that the number of electron events leaking into the proton branch is negligible compared to the 1% cosmic background. **b**, Proton recoil spectrum. Valid neutron events are shown in histogram format. For comparison, we also show our detector's simulated<sup>20,21</sup> responses to 1.45 MeV, 2.45 MeV and 3.45 MeV centre-of-mass boosted neutrons.



**Figure 4** Neutron time-of-flight measurement. **a**, A deuteron is shown striking a thin disk of deuterated plastic scintillator, where it fuses with another deuteron, producing an 820-keV  $^3\text{He}$  and a 2.45-MeV neutron. The  $\alpha$ -particle promptly scintillates in the plastic, recorded by a photomultiplier tube coupled to the glass UHV viewport through a silicone optical pad. The neutron, on the other hand, leaves the vacuum chamber, and is shown detected via proton recoil in the liquid scintillator. **b**, Experiment geometry. **c**, Simultaneously captured PMT traces, demonstrating an  $\alpha$ -particle–neutron coincidence. The plastic scintillator trace, shown in the upper panel, has a large  $\alpha$ -particle hit at  $t = 0$  ns, whereas the smaller hits are incident deuterons that stopped in the plastic

but did not fuse. The liquid scintillator trace, shown in the lower panel, has a proton hit at  $t = 6$  ns. **d**, Time-of-flight results. The distribution of neutron flight times is shown in the upper histogram. As the neutron emission and detection volumes are finite and relatively closely spaced, we observe a range of flight times. The Monte Carlo flight time distribution, including a constant term to account for background, is shown fitted. The peak in the distribution roughly corresponds with the 5.6 ns it takes a 2.45-MeV neutron moving with a velocity of  $0.07c$  (where  $c$  is the speed of light) to travel 12 cm. The relative timing offset between the two PMTs was calibrated using back-to-back 511-keV  $\gamma$ -rays from a  $^{22}\text{Na}$  source, as shown in the lower histogram.

neutron flux to the neutron flux expected from the ion beam striking the  $\text{ErD}_2$  target. At the time of peak neutron flux, the ion current was 4.2 nA and the accelerating potential, inferred from the bremsstrahlung endpoint, was 115 kV. Using tabulated stopping powers<sup>13</sup> and fusion cross-sections<sup>14</sup>, we calculate a neutron flux of  $900 \text{ neutrons s}^{-1}$ . This is a slight overestimate, because part of the ion beam struck outside the target and there was an oxide layer on the target.

In Fig. 4, we present our neutron time-of-flight measurement. Using deuterated plastic scintillator (BC-436) as both a deuterated target, and as a scintillation material, allowed us to pinpoint individual fusion events. The scintillator was mounted inside the chamber against a glass ultrahigh-vacuum (UHV) viewport, through which a Hamamatsu H1949-50 PMT was coupled via a silicone optical pad. The side of the scintillator facing the beam had a 50-nm layer of evaporated aluminium and was connected to the picoammeter. The aluminium prevented the target from charging up, allowed for a reliable beam current measurement, and helped screen out stray light originating from within the chamber. To minimize background hits, yet still collect valid coincidences, we used a reduced deuterium pressure and a reduced heating rate so that the ion current was around 10 pA. Running at this low level permitted prolonged runs. For example, the data shown in Fig. 4d were taken from a single heating cycle lasting over eight hours.

We have shown that small (about centimetre-sized) pyroelectric crystals can produce ion beams (see also Supplementary Fig. 1 and Supplementary Movie 2) of sufficient energy and current to drive nuclear fusion. We anticipate increasing the field ionization current by using a larger tip, or tip array, and by operating at cryogenic temperatures. With these enhancements, and in addition using a tritiated target, we believe that the reported signal could be scaled beyond  $10^6 \text{ neutrons s}^{-1}$ . Pyroelectric crystals may also have applications in electrostatic fusion devices<sup>15</sup>, such as the Farnsworth fusor<sup>16–18</sup>, and as microthrusters in miniature spacecraft<sup>19</sup>. □

Received 14 December 2004; accepted 30 March 2005; doi:10.1038/nature03575.

- Shimomura, Y. & Spears, W. Review of the ITER project. *IEEE Trans. Appl. Supercond.* **14**, 1369–1375 (2004).
- Miller, G. H., Moses, E. I. & Wuest, C. R. The National Ignition Facility: enabling fusion ignition for the 21st century. *Nucl. Fusion* **44**, S228–S238 (2004).
- Ditmire, T. *et al.* Nuclear fusion from explosions of femtosecond laser-heated deuterium clusters. *Nature* **398**, 489–492 (1999).

- Pretzler, G. *et al.* Neutron production by 200 mJ ultrashort laser pulses. *Phys. Rev. E* **58**, 1165–1168 (1998).
- Taubes, G. *Bad Science: The Short Life and Weird Times of Cold Fusion* (Random House, New York, 1993).
- Taleyarkhan, R. P. *et al.* Evidence for nuclear emissions during acoustic cavitations. *Science* **295**, 1868–1873 (2002).
- Salmarsh, M. J. & Shapira, D. Questions regarding nuclear emissions in cavitation experiments. *Science* **297**, 1603 (2002).
- Glass, A. M. Dielectric, thermal, and pyroelectric properties of ferroelectric  $\text{LiTaO}_3$ . *Phys. Rev.* **172**, 564–571 (1968).
- Rosenblum, B., Braunlich, P. & Carrico, J. P. Thermally stimulated field emission from pyroelectric  $\text{LiNbO}_3$ . *Appl. Phys. Lett.* **25**, 17–19 (1974).
- Riege, H. Electron emission from ferroelectrics—a review. *Nucl. Instrum. Methods A* **340**, 80–89 (1994).
- Brownridge, J. D., Shafroth, S. M., Trott, D. W., Stoner, B. R. & Hooke, W. M. Observation of multiple nearly monoenergetic electron production by heated pyroelectric crystals in ambient gas. *Appl. Phys. Lett.* **78**, 1158–1159 (2001).
- Rosenman, G., Shur, D., Krasik, Ya. E. & Dunaevsky, A. Electron emission from ferroelectrics. *J. Appl. Phys.* **88**, 6109–6161 (2000).
- Ziegler, J. F. The stopping of energetic ions in solids. *Nucl. Instrum. Methods* **168**, 17–24 (1980).
- Bosch, H. S. & Hale, G. M. Improved formulas for fusion cross-sections and thermal reactivities. *Nucl. Fusion* **32**, 611–631 (1992).
- Nevins, W. M. Can inertial electrostatic confinement work beyond the ion-ion collisional time scale? *Phys. Plasmas* **2**, 3804–3819 (1995).
- Farnsworth, P. T. Electric discharge device for producing interactions between nuclei. US Patent No. 3258402 (1966).
- Farnsworth, P. T. Method and apparatus for producing nuclear-fusion reactions. US Patent No. 3386883 (1968).
- Hirsch, R. L. Inertial-electrostatic confinement of ionized fusion gases. *J. Appl. Phys.* **38**, 4522–4534 (1967).
- Mitterauer, J. Micropropulsion for small spacecraft: a new challenge for field effect electric propulsion and microstructured liquid metal ion sources. *Surf. Interface Anal.* **36**, 380–386 (2004).
- Agostinelli, S. *et al.* GEANT4 — a simulation toolkit. *Nucl. Instrum. Methods A* **506**, 250–303 (2003).
- Verbinski, V. V. *et al.* Calibration of an organic scintillator for neutron spectrometry. *Nucl. Instrum. Methods* **65**, 8–25 (1968).

**Supplementary Information** accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** The neutron detector was built with funds from DARPA. Funding for various stages of this project was provided by the NSF, ONR and DARPA. We thank W. Wright and K. O'Doherty for demonstrations of electron and ion emission from pyroelectrics, respectively; H. Lockart for machine workshop expertise; K. O'Doherty for evaporating the 50-nm film onto the plastic scintillator; R. Cousins for recommending the use of liquid scintillator for detection and pulse-shape identification of neutrons, and for overseeing the design and construction of a prototype detector using waveform digitization; and T. Venhaus, W. Harbin and J. Hoffer of LANL (ESA-TSE group) for supplying the deuterated target. S. P. thanks A. Erbil for bringing the phenomenon of ferroelectric emission to his attention, along with ref. 10.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to B.N. ([naranjo@physics.ucla.edu](mailto:naranjo@physics.ucla.edu)).

# Increased productivity in the subantarctic ocean during Heinrich events

Julian P. Sachs<sup>1</sup> & Robert F. Anderson<sup>2</sup>

<sup>1</sup>Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Room E34-254, Cambridge, Massachusetts 02139, USA

<sup>2</sup>Lamont-Doherty Earth Observatory of Columbia University, PO Box 1000, Palisades, New York 10964, USA

Massive iceberg discharges from the Northern Hemisphere ice sheets, 'Heinrich events', coincided with the coldest periods of the last ice age<sup>1</sup>. There is widespread evidence for Heinrich events and their profound impact on the climate and circulation of the North Atlantic Ocean, but their influence beyond that region remains uncertain<sup>1</sup>. Here we use a combination of molecular fingerprints of algal productivity and radioisotope tracers of sedimentation to document eight periods of increased productivity in the subpolar Southern Ocean during the past 70,000 years that occurred within 1,000–2,000 years of a Northern Hemisphere Heinrich event. We discuss possible causes for such a link, including increased supply of iron from upwelling and increased stratification during the growing season, which imply an alteration of the global ocean circulation during Heinrich events. The mechanisms linking North Atlantic iceberg discharges with subantarctic productivity remain unclear at this point. We suggest that understanding how the Southern Ocean was altered during these extreme climate perturbations is critical to understanding the role of the ocean in climate change.

Ice-age cycles are paced by periodic changes in Earth's orbital geometry that modulate the amount of solar radiation reaching Northern Hemisphere ice sheets<sup>2</sup>. For reasons that remain unclear, the climate system is susceptible to sudden shifts during glacial intervals, when continental ice sheets are large, sea level and atmospheric greenhouse gas concentrations are low, and sea ice in both hemispheres is extensive. Instability arising from ice-sheet dynamics, stochastic resonance in the climate system or some yet-to-be-identified forcing mechanism caused massive surges of icebergs from Northern Hemisphere ice sheets to the North Atlantic every ~5–10 kyr during glacial intervals of the Pleistocene epoch that coincided with the harshest conditions of the glacial period in Europe<sup>1</sup>. Large debris fields spanning the North Atlantic sea floor are the remnants of so-called Heinrich events. Gravel in marine sediments thousands of kilometres from land must have been delivered by icebergs whose debris-laden undersides once scoured the continents.

Eight such episodes (Heinrich events and the Younger Dryas event) occurred between 10 and 70 kyr ago<sup>1,3,4</sup>. Each resulted in freshening and cooling of surface waters of the North Atlantic<sup>3,1</sup>. Synchronous changes in the chemistry or biology of sediment cores from the southwest equatorial Atlantic<sup>5</sup>, the southeast subtropical Atlantic<sup>6</sup> and the Arabian Sea, and in a stalagmite from southeastern China, argue for a widespread impact of Heinrich events on climate<sup>1</sup>.

Although some data<sup>1,7</sup> and models<sup>8,9</sup> suggest that flooding the North Atlantic with icebergs arrested production of North Atlantic Deep Water, thus causing a decrease in the poleward flux of heat to Europe, other data<sup>10</sup> and models<sup>11</sup> suggest otherwise. That current models and data do not agree on the nature, sign or magnitude of change in North Atlantic thermohaline circulation caused by the most extreme climate events of the past 100,000 years

illustrates the need for further study of the coupling between ocean circulation and abrupt climate change.

The Southern Ocean plays a central role in the global thermohaline circulation. Most deep-water production occurs there, as does about half of the return (upwelling) flux of deep water to the surface<sup>12</sup>. Although prominent warm episodes in Antarctica coincided with six of the eight Heinrich events (including the Younger Dryas)<sup>13</sup>, almost nothing is known about this southern branch of the global thermohaline circulation during Heinrich events. Here we present the first evidence of major perturbations in the Southern Ocean during Heinrich events, by demonstrating that algal productivity soared in subantarctic waters of the southwest Pacific and southeast Atlantic oceans at those times.

Algal productivity in the southwest Pacific east of New Zealand was inferred from the sedimentary concentration of alkenones, C<sub>37</sub> methyl ketones unique to coccolithophorid algae, and brassicasterol (24-methylcholesta-5,22E-dien-3 $\beta$ -ol), a marker for diatoms<sup>14</sup>. Alkenones were measured every 2 cm in the upper 10.39 m, and brassicasterol and *n*-octacosanol were measured every 2–4 cm between 4.74 and 10.01 m, of core MD97-2120, retrieved from the Chatham rise, east of the South Island of New Zealand, at 45° 32.06' S, 174° 55.85' E, in 1,210 m of water. The site underlies subantarctic water and has probably remained south of the Subtropical Convergence as a result of bathymetric control provided by the Chatham rise of the position of the Subtropical Front<sup>15</sup>. A chronology for the core was developed by Pahnke *et al.*<sup>16</sup> from ten radiocarbon dates between 0 and 36 kyr ago, the <sup>14</sup>C-dated Kawakawa tephra layer at 26.2 kyr ago, and by graphical correlation of a benthic foraminiferal oxygen isotope record in core MD97-2120 with a reference record tied to the Greenland GISP2 ice-core chronology. Age control points are shown at the top of Fig. 1.

Alkenone and brassicasterol concentrations exhibit distinct concentration maxima coinciding with each Heinrich event (or the Younger Dryas) within the uncertainty of the age model (Fig. 1a). Uncertainty in the age of the biomarker maxima—especially the most prominent events 65–35 kyr ago that are too old to be dated by radiocarbon—together with the uncertainties in the actual ages of the Heinrich events, makes it impossible to determine whether the biomarker maxima lead, lag or are coincident with Heinrich events. Nevertheless, even with dating uncertainties of several thousand years before 35 kyr ago, the number and timing of subantarctic biomarker maxima are remarkably similar to the pattern of Heinrich events. Furthermore, there are about half as many biomarker peaks as there are interstadial (Dansgaard-Oeschger) events in Greenland ice during the studied interval, making a causal relationship with D–O events unlikely. We thus interpret the similar patterns of biomarker maxima and Heinrich events to indicate a mechanistic relationship between them.

Increased concentrations of algal biomarkers in subantarctic sediments near the time of Heinrich events were not limited to the Chatham rise. Alkenone concentration maxima occurred with near synchrony half way around the globe in sediments of the Cape basin, southeast Atlantic Ocean (core TN057-21-PC2, from 41° 08' S, 7° 49' E and 4,981 m water depth), during the period 65–35 kyr ago, when Heinrich events H4–H6 occurred (Fig. 1d). The similar number, timing and magnitude of lipid biomarker concentration changes at the two sites 65–35 kyr ago (Fig. 1a,d) suggests that the events may have had a common cause.

Understanding what caused the biomarker maxima will provide important clues about conditions in the Southern Ocean during Heinrich events. Three factors influence algal biomarker concentrations in sediments: (1) their production rate in overlying surface water, (2) their preservation rate in the water column and sediments, and (3) their dilution by other sedimentary components.

The influence of changing preservation rates was evaluated first by measuring the concentration of the terrestrial plant biomarker, *n*-octacosanol (C<sub>28</sub>-*n*-alcohol), and comparing its variation in

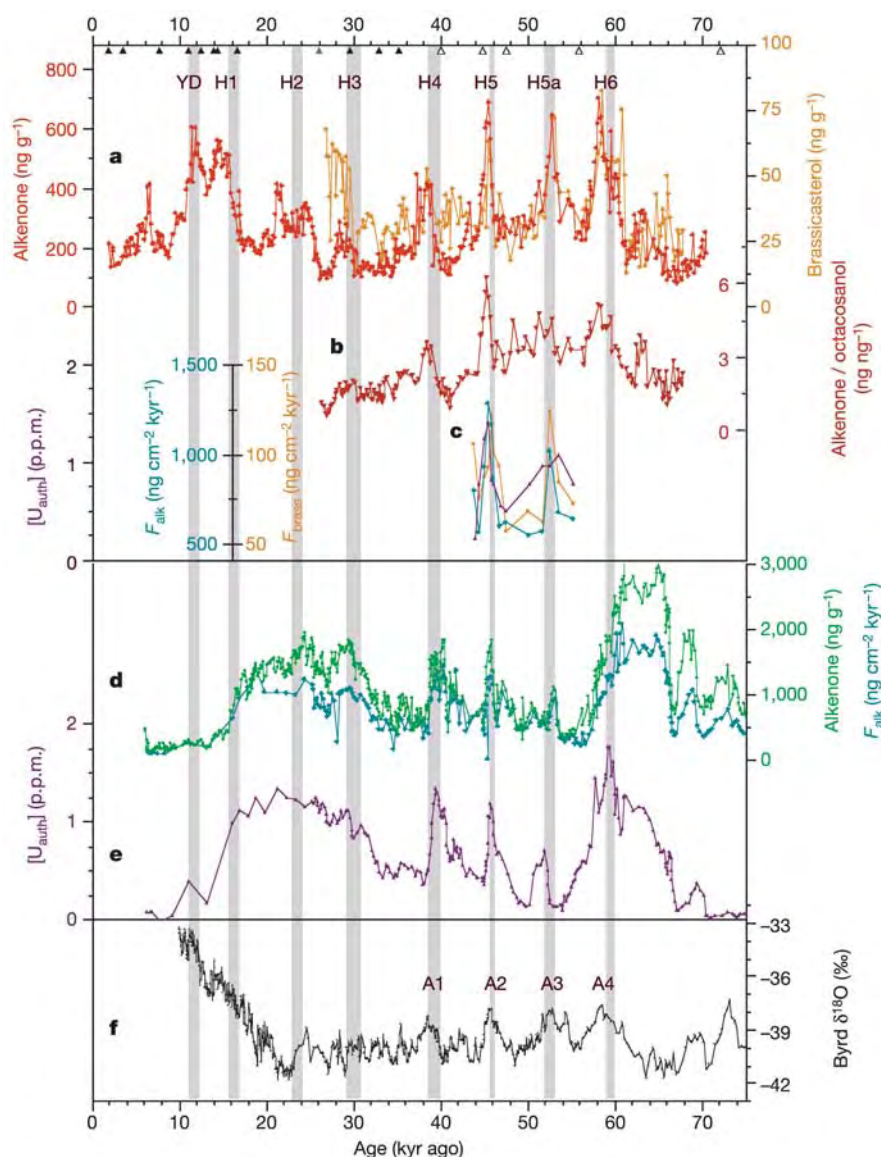


MD97-2120 sediment to that of alkenones (Fig. 1b). A maximum in the ratio of the marine algal lipid to the terrestrial plant lipid occurred in association with each alkenone maximum 65–27 kyr ago, implying that diagenetic reactions, which are expected to influence the two lipids similarly, were not the primary cause of the alkenone concentration maxima.

A second argument against changing preservation rates being the main driver of alkenone concentration changes comes from the hydrography at the two sites where water depth, sediment accumulation rate and bottom-water chemistry differ substantially.

At depths of 1,210 m, the Pacific site lies within oxygen-rich Antarctic Intermediate Water. At depths of 4,981 m, the Atlantic site lies within oxygen-depleted Circumpolar Deep Water. It is difficult to imagine a mechanism for improving alkenone preservation at both sites simultaneously by changing ocean circulation and related changes in the concentration of dissolved oxygen in waters overlying the core sites.

The influence of dilution by other sedimentary components on algal lipid concentrations was evaluated by computing  $^{230}\text{Th}$ -normalized biomarker fluxes<sup>17,18</sup>. A maximum of  $^{230}\text{Th}$ -normalized



**Figure 1** Subantarctic productivity changes during Heinrich events. **a**, Alkenone (red) and brassicasterol (orange) concentrations in Chatham rise core MD97-2120. Covariation of these biomarkers for coccolithophorids and diatoms, respectively, indicates that algal productivity increased during Heinrich events. **b**, Ratio of alkenones to *n*-octacosanol in core MD97-2120. An excess of algal lipids (alkenones) relative to terrestrial plant lipids (*n*-octacosanol) in H3–H6 implies that lipid preservation was not the primary cause of the algal biomarker concentration increases. **c**,  $^{230}\text{Th}$ -normalized flux of alkenones ( $F_{\text{alk}}$ ; blue) and brassicasterol ( $F_{\text{brass}}$ ; orange), and the concentration of  $U_{\text{auth}}$  (purple) in Chatham rise core MD97-2120. **d**, Alkenone concentration (green) and  $^{230}\text{Th}$ -normalized alkenone flux (blue) in Cape basin (southeast Atlantic) core TN057-21-PC2 (refs 17, 18). **e**, The concentration of authigenic uranium in Cape basin core TN057-21-PC2 (refs 17, 18). **f**,  $\delta^{18}\text{O}$  variations, a proxy for air temperature, in the Byrd, Antarctica, ice core<sup>13</sup>. Prominent Antarctic temperature maxima A1–A4 (ref. 13) coincide with subantarctic

productivity maxima and H4–H6. Age control points for MD97-2120 are shown at the top of panel **a**. Black triangles correspond to  $^{14}\text{C}$  dates on the planktonic foraminifera *Globigerina bulloides*, the grey triangle to the Kawakawa tephra, and open triangles to oxygen isotopic tie points to core MD95-2042 in the North Atlantic<sup>16</sup>. Age control for TN057-21-PC2 was provided by graphical comparison of magnetic intensity variations in the sediment with cosmogenic isotope production changes in Greenland ice<sup>30</sup>. Calendar ages of the Heinrich events and the Younger Dryas (YD) are indicated by vertical grey bars, the thickness of which span their duration and age range from recent studies<sup>3,4</sup>. Concentrations of alkenones, *n*-octacosanol and brassicasterol were determined by gas chromatography with flame-ionization detection and co-injected standards, following identification by gas chromatography-mass spectrometry<sup>18</sup>. Concentrations of U and Th isotopes were measured by isotope dilution inductively coupled plasma mass spectrometry<sup>18</sup>.

alkenone flux occurred at every maximum of alkenone concentration in sediments of the Cape basin (Fig. 1d) and at the two alkenone maxima in which we measured U and Th in sediments of the Chatham rise (Fig. 1c). Thus dilution by other sedimentary components did not cause the algal lipid concentration changes we observe. For this reason, and those discussed below, the most likely cause of increased alkenone concentrations in sediments of the Chatham rise and the Cape basin during Heinrich events was increased algal productivity.

Covariation of brassicasterol concentrations with alkenone concentrations in Chatham rise sediment 65–35 kyr ago (Fig. 1a) implies that increases in algal productivity were not limited to coccolithophorids, but probably included diatoms as well. Brassicasterol is often used as a molecular indicator for the presence of diatoms, although it is also produced by some prymnesiophytes<sup>14</sup>. High <sup>230</sup>Th-normalized fluxes of brassicasterol 46 and 52 kyr ago (Fig. 1c) supports the notion of higher algal productivity in subantarctic waters at those times.

Additional support for higher algal production comes from the concentration of authigenic uranium ( $U_{\text{auth}}$ ) in Cape basin (Fig. 1e) and Chatham rise (Fig. 1c) sediment, which increased in concert with alkenone concentrations during at least six of eight Heinrich events (Younger Dryas and H3–H6) in the Cape basin and the two Heinrich events (H5 and H5a) analysed at the Chatham rise. Both organic flux and bottom-water oxygen concentration influence  $U_{\text{auth}}$  concentrations in sediments of the Southern Ocean<sup>19</sup>. Although the two parameters are not independent—that is, remineralization of organic matter uses oxygen—previous studies have attributed increased  $U_{\text{auth}}$  in Southern Ocean sediments to increased organic flux<sup>19</sup>. Furthermore, as noted for algal lipid preservation, reduced ventilation and resultant depletion of bottom-water oxygen in response to altered ocean circulation would not be expected to occur concurrently in mid-depths (1,210 m) of the southwest Pacific and abyssal depths (4,981 m) of the southeast Atlantic.  $U_{\text{auth}}$  did not increase appreciably during H1 or H2 in the Cape basin record, perhaps because  $U_{\text{auth}}$  was limited by uranium diffusion from sea water when export production was high during marine isotope stage 2 (MIS2).

Lastly, we note that the <sup>230</sup>Th-normalized mass flux of sediment to the Chatham rise was nearly constant ( $2.2 \pm 0.4 \text{ g cm}^{-2} \text{ kyr}^{-1}$ ,  $n = 13$ ) 43–57 kyr ago, implying that changes in sediment delivery to the site were not associated with the biomarker changes. By eliminating other possible causes, we conclude that an increase in the export flux of organic matter from subantarctic surface waters south of the Chatham rise and in the Cape basin occurred near the time of Heinrich events in the Northern Hemisphere.

Evaluating the factors that may have caused increased biological productivity puts constraints on conditions in the Southern Ocean during Heinrich events. Iron limits algal productivity in the Southern Ocean<sup>20</sup>. Iron-addition experiments and molecular markers for iron stress in phytoplankton south of the Chatham rise confirm that Fe limits algal growth equatorward to the Subtropical Front<sup>21</sup>. If coccolithophorid and diatom productivity increased simultaneously, as implied by the alkenone and brassicasterol data (Fig. 1a), and the export flux of organic matter increased, as implied by the authigenic uranium activity and the algal lipid fluxes (Fig. 1a–c), increased iron availability may have been the cause.

Unlike most of the ocean, in the Southern Ocean iron comes primarily from upwelling of deep water rather than from direct input of continental dust<sup>22,23</sup>. Whether this was the case during glacial times is unknown, but dust proxies in Antarctic ice cores indicate diminished dust deposition during the productivity events<sup>24</sup>. If dust-borne iron fluxes to the Southern Ocean did not increase during Heinrich events, the most likely source of additional iron to subantarctic surface water was an increased upwelling flux of iron-enriched deep water.

Phytoplankton growth in the Southern Ocean is also limited by

light, at least at certain times and in certain places<sup>25,26</sup>. Factors leading to increased exposure to light by phytoplankton, such as increased stratification of surface waters<sup>27</sup>, may also have contributed to increased productivity during Heinrich events. A compelling association exists between warm episodes A1–A4 in Antarctica<sup>13</sup> and the most prominent four productivity maxima (at 58.5, 53, 46, and 38.5 kyr ago) (Fig. 1f). Greater stratification of Southern Ocean surface water may have occurred at these times, either by radiative heating, decreased wind stress (owing to the thermal wind relationship and supported by wind proxies in Antarctic ice<sup>24</sup>), or by increased freshwater supply associated with melting of the Antarctic Ice Sheet<sup>28</sup>.

The largest increases in subantarctic productivity occurred 35–65 kyr ago during H4–H6 (Fig. 1a–e) and coincided with the most prominent CO<sub>2</sub> (ref. 29) and Antarctic temperature<sup>13</sup> (Fig. 1f) maxima of the last glacial period. Why did productivity, CO<sub>2</sub> and temperature rise only modestly, if at all, during the Younger Dryas and H1–H3? One possibility is that conditions in the North Atlantic during Heinrich events were more similar to those during the height of the Last Glacial Maximum period 12–27 kyr ago (MIS2)—that is, colder, windier, diminished North Atlantic Deep Water production—than during the preceding warmer period 27–58 kyr ago (MIS3). Heinrich events that occurred during MIS3 may have therefore caused a larger (relative) perturbation to the existing ocean and atmospheric circulation than those that occurred during MIS2.

Existing data do not permit an unequivocal distinction between upwelling and stratification as the principal factor responsible for increased subantarctic productivity associated with Heinrich events. Neither are the two factors mutually exclusive, as increased summer stratification may have coincided with times of increased upwelling. The value of these new records is that they provide the first evidence for prominent alterations of the Southern Ocean during Heinrich events. As such, they define targets for future research. For example, future studies should explore evidence for increased abundance of icebergs and reduced surface salinity to test the possibility that pulses of ice (freshwater) from Antarctica increased stratification at times of increased productivity.

Understanding how the Southern Ocean was altered during extreme climate perturbations is critical to understanding the role of ocean circulation in climate change. A glaring lack of palaeoclimate data hinders this effort. The observation of subantarctic productivity maxima that correspond closely with Heinrich events (in number and timing) and with Antarctic warm events (in relative magnitude and timing) but poorly with D–O events, gets us a step closer towards understanding the mechanisms of abrupt climate change. □

Received 20 December 2004; accepted 8 March 2005; doi:10.1038/nature03544.

- Hemming, S. R. Heinrich events: massive late Pleistocene detritus layers of the North Atlantic and their global climate imprint. *Rev. Geophys.* **42**, doi:10.1029/2003RG000128 (2004).
- Hays, J. D., Imbrie, J. & Shackleton, N. J. Variations in the Earth's orbit: pacemaker of the ice ages. *Science* **194**, 1121–1132 (1976).
- van Kreveld, S. *et al.* Potential links between surging ice sheets, circulation changes, and the Dansgaard-Oeschger cycles in the Irminger Sea, 60–18 kyr. *Paleoceanography* **15**, 425–442 (2000).
- Rashid, H., Hesse, R. & Piper, D. J. W. Evidence for an additional Heinrich event between H5 and H6. *Paleoceanography* **18**, 1077, doi:10.1029/2003PA000913 (2003).
- Arz, H. W., Pätzold, J. & Wefer, G. Correlated millennial-scale changes in surface hydrography and terrigenous sediment yield inferred from last-glacial marine deposits off northeastern Brazil. *Quat. Res.* **50**, 157–166 (1998).
- Little, M. G. *et al.* Trade wind forcing of upwelling, seasonality, and Heinrich events as a response to sub-Milankovitch climate variability. *Paleoceanography* **12**, 568–576 (1997).
- McManus, J. F., Francois, R., Gherardi, J.-M., Keigwin, L. D. & Brown-Leger, S. Collapse and rapid resumption of Atlantic meridional circulation linked to deglacial climate changes. *Nature* **428**, 834–837 (2004).
- Rind, D. *et al.* Effects of meltwater in the GISS coupled atmosphere-ocean model 2. A bipolar seesaw in Atlantic Deep Water production. *J. Geophys. Res.* **106**, 27355–27365 (2001).
- Rahmstorf, S. Ocean circulation and climate during the past 120,000 years. *Nature* **419**, 207–214 (2002).
- Piotrowski, A. M., Goldstein, S. L., Hemming, S. R. & Fairbanks, R. G. Intensification and variability of ocean thermohaline circulation through the last deglaciation. *Earth Planet. Sci. Lett.* **225**, 205–220 (2004).
- Nilsson, J., Brostrom, G. & Walin, G. The thermohaline circulation and vertical mixing: Does weaker

- density stratification give stronger overturning? *J. Phys. Oceanogr.* **33**, 2781–2795 (2003).
12. Webb, D. J. & Sugimotohara, N. Vertical mixing in the ocean. *Nature* **409**, 37 (2001).
  13. Blunier, T. & Brook, E. J. Timing of millennial-scale climate change in Antarctica and Greenland during the last glacial period. *Science* **291**, 109–112 (2001).
  14. Volkman, J. K. *et al.* Microalgal biomarkers: A review of recent research developments. *Org. Geochem.* **29**, 1163–1179 (1998).
  15. Weaver, P. P. E., Carter, L. & Neil, H. L. Response of surface water masses and circulation to late Quaternary climate change east of New Zealand. *Paleoceanography* **13**, 70–83 (1998).
  16. Pahnke, K., Zahn, R., Elderfield, H. & Schulz, M. 340,000-year centennial-scale marine record of Southern Hemisphere climatic oscillation. *Science* **301**, 948–952 (2003).
  17. Sachs, J. P., Anderson, R. F. & Lehman, S. J. Glacial surface temperatures of the southeast Atlantic Ocean. *Science* **293**, 2077–2079 (2001).
  18. Sachs, J. P. & Anderson, R. F. Fidelity of alkenone paleotemperature reconstructions in southern Cape basin sediment drifts. *Paleoceanography* **18**, 1082, doi:10.1029/2002PA000862 (2003).
  19. Chase, Z., Anderson, R. F. & Fleisher, M. Q. Evidence from authigenic uranium for increased productivity of the glacial Subantarctic Ocean. *Paleoceanography* **16**, 468–478 (2001).
  20. Martin, J. H. Glacial-interglacial CO<sub>2</sub> change: the iron hypothesis. *Paleoceanography* **5**, 1–13 (1990).
  21. Boyd, P., LaRoche, J., Gall, M., Frew, R. & McKay, R. M. L. Role of iron, light, and silicate in controlling algal biomass in subantarctic waters SE of New Zealand. *J. Geophys. Res.* **104**, 13395–13408 (1999).
  22. Watson, A. J. & Lefevre, N. The sensitivity of atmospheric CO<sub>2</sub> concentrations to input of iron to the oceans. *Tellus* **51B**, 453–460 (1999).
  23. Lefevre, N. & Watson, A. J. Modeling the geochemical cycle of iron in the oceans and its impact on atmospheric CO<sub>2</sub> concentrations. *Glob. Biogeochem. Cycles* **13**, 727–736 (1999).
  24. Steig, E. J. *et al.* Wisconsinan and Holocene climate history from an ice core at Taylor Dome, western Ross Embayment, Antarctica. *Geogr. Ann.* **82A**, 213–235 (2000).
  25. Abbott, M. R., Richman, J. G., Letellier, R. M. & Bartlett, J. S. The spring bloom in the Antarctic Polar Frontal Zone as observed from a mesoscale array of bio-optical sensors. *Deep-Sea Res.* **47**, 3285–3314 (2000).
  26. Mitchell, B. G., Brody, E. A., Holm-Hansen, O., McClain, C. & Bishop, J. Light limitation of phytoplankton biomass and macronutrient utilization in the Southern Ocean. *Limnol. Oceanogr.* **36**, 1662–1677 (1991).
  27. Smith, W. O. Jr & Nelson, D. M. Phytoplankton bloom produced by a receding ice edge in the Ross Sea: spatial coherence with the density field. *Science* **227**, 163–167 (1985).
  28. Rohling, E. J., Marsh, R., Wells, N. C., Siddall, M. & Edwards, N. R. Similar meltwater contributions to glacial sea level changes from Antarctic and northern ice sheets. *Nature* **430**, 1016–1021 (2004).
  29. Indermühle, A., Monnin, E., Stauffer, B., Stocker, T. F. & Wahlen, M. Atmospheric CO<sub>2</sub> concentration from 60 to 20 kyr BP from the Taylor Dome ice core, Antarctica. *Geophys. Res. Lett.* **27**, 735–738 (2000).
  30. Stoner, J. S., Channell, J. E. T., Hillaire-Marcel, C. & Kissel, C. Geomagnetic paleointensity and environmental record from Labrador Sea core MD95–2024: global marine sediment and ice core chronostratigraphy for the last 110 kyr. *Earth Planet. Sci. Lett.* **183**, 161–177 (2000).

**Acknowledgements** Discussions with C. Wunsch, J. Marshall, M. Follows, E. Boyle and P. Parekh contributed to this manuscript. Samples from core MD97–2120 were provided by K. Pahnke and R. Zahn. The Lamont-Doherty Earth Observatory core repository provided samples from core TN057–21-PC2. D. Dryer, M. Fleisher, Y. Chang and M. Bryan assisted with laboratory analyses. Funding for J.P.S. was from the Gary Comer Foundation, the Jephtha H. and Emily V. Wade Award for Research, and a Henry L. and Grace Doherty Professorship. Funding for R.F.A. was from a grants/cooperative agreement from the National Oceanic and Atmospheric Administration.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to J.P.S. (jsachs@mit.edu).

## Electronic tagging and population structure of Atlantic bluefin tuna

Barbara A. Block<sup>1</sup>, Steven L. H. Teo<sup>1\*</sup>, Andreas Walli<sup>1\*</sup>, Andre Boustany<sup>1\*</sup>, Michael J. W. Stokesbury<sup>1,3</sup>, Charles J. Farwell<sup>2</sup>, Kevin C. Weng<sup>1</sup>, Heidi Dewar<sup>1</sup> & Thomas D. Williams<sup>2</sup>

<sup>1</sup>Tuna Research and Conservation Center, Stanford University, Hopkins Marine Station, Pacific Grove, California 93950, USA

<sup>2</sup>Monterey Bay Aquarium, 886 Cannery Row, Monterey, California 93940, USA

<sup>3</sup>Dalhousie University, Biology Department, Halifax, Nova Scotia, B3H 4J1 Canada

\* These authors contributed equally to this work

**Electronic tags that archive or transmit stored data to satellites have advanced the mapping of habitats used by highly migratory fish in pelagic ecosystems<sup>1–6</sup>. Here we report on the electronic tagging of 772 Atlantic bluefin tuna in the western Atlantic Ocean in an effort to identify population structure. Reporting electronic**

**tags provided accurate location data<sup>7–9</sup> that show the extensive migrations of individual fish ( $n = 330$ ). Geoposition data delineate two populations, one using spawning grounds in the Gulf of Mexico and another from the Mediterranean Sea. Transatlantic movements of western-tagged bluefin tuna reveal site fidelity to known spawning areas in the Mediterranean Sea. Bluefin tuna that occupy western spawning grounds move to central and eastern Atlantic foraging grounds. Our results are consistent with two populations of bluefin tuna with distinct spawning areas that overlap on North Atlantic foraging grounds. Electronic tagging locations, when combined with US pelagic longline observer and logbook catch data, identify hot spots for spawning bluefin tuna in the northern slope waters of the Gulf of Mexico. Restrictions on the time and area where longlining occurs would reduce incidental catch mortalities on western spawning grounds.**

Giant bluefin tuna are the largest members of the family Scombridae, attaining body sizes of more than 650 kg (refs 10, 11). They are unique among teleosts for their endothermic capacity and cardiovascular physiology<sup>12,13</sup>. These traits underlie their capacity to exploit environments ranging from subarctic feeding grounds to subtropical spawning areas. Top pelagic predators such as bluefin tuna are in precipitous decline globally because of overexploitation<sup>14</sup>. The International Commission for the Conservation of Atlantic Tunas (ICCAT) manages Atlantic bluefin tuna as distinct western and eastern stocks separated by a management boundary at the 45° W meridian<sup>10,11</sup>. The spawning stock biomass of western Atlantic bluefin tuna has decreased by 80% or more since 1970 (refs 10, 11). A 20-year rebuilding plan was enacted in the early 1980s in the western Atlantic<sup>10</sup>. The most recent assessment indicates that the western stock continues to decline<sup>11</sup>, yet mortality throughout the North Atlantic remains high. Key questions remain on the biology of this species. Establishing the location and timing of reproduction, the mean age of maturity, spawning site fidelity, the ontogeny of movement patterns and the influence of climate variability on movements will improve stock assessments and subsequent management<sup>15</sup>. Here we report the spatio-temporal distributions of Atlantic bluefin tuna determined with electronic tags, discriminate two potential spawning populations, and record spawning site fidelity to the Mediterranean Sea.

We deployed 499 implantable archival tags and 273 pop-up satellite (PAT) tags on bluefin tuna in the western Atlantic (Supplementary Information)<sup>2,3,6</sup>. To date, 86 archival-tagged bluefin tuna have been recaptured; 54 in the west Atlantic, 9 in the east Atlantic and 23 in the Mediterranean Sea. Twelve PAT-tagged fish were recaptured and 237 PAT tags transmitted data to Argos satellites after 2 to 251 days after tagging (Table 1). Individual tracks of 2 to 1,623 days have been obtained.

Our database comprises 13,372 positions obtained from 330 bluefin tuna that carried electronic tags from 1996 to 2004 (Fig. 1, Table 1). Geoposition data include the following: Doppler-based Argos endpoint positions calculated for PAT tags ( $n = 237$ )<sup>9</sup>; geolocation estimates for archival ( $n = 5,171$ ) and PAT tags ( $n = 7,536$ ), using light level and sea surface temperature (SST) to estimate longitude and latitude, respectively<sup>9</sup>; Global Positioning System deployment locations for recovered archival and reporting PAT tags ( $n = 330$ ); and recapture locations from recovered archival and PAT tags ( $n = 98$ ). The distribution of these positions across the North Atlantic Ocean indicates that the western and eastern management units are strongly linked with overlapping ranges.

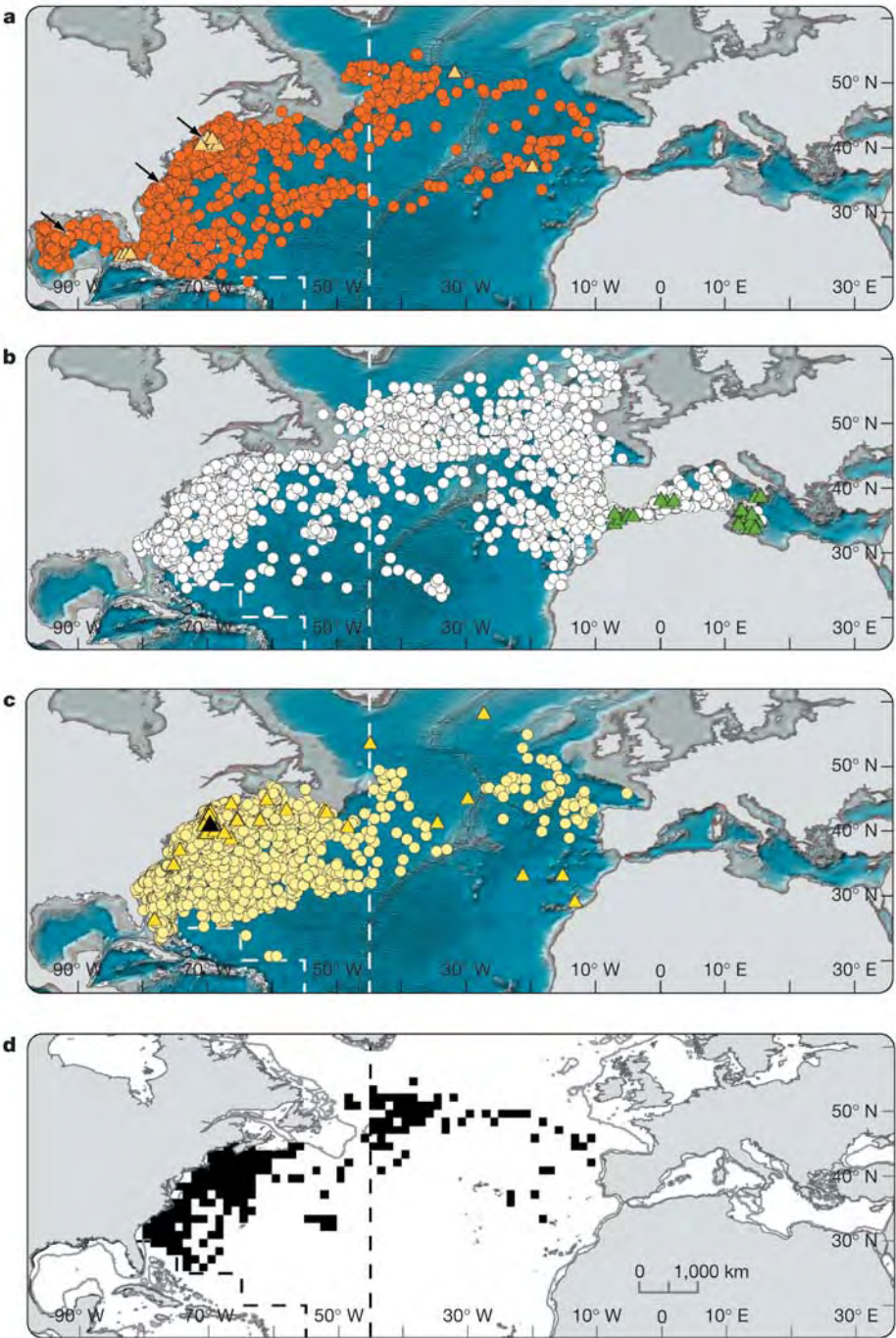
The electronic tagging data reveal two populations of Atlantic bluefin tuna that overlap on North Atlantic Ocean foraging grounds and sort to independent spawning areas located primarily in the Gulf of Mexico (GOM) and Mediterranean Sea (Fig. 1). A bluefin tuna was assigned to the western Atlantic spawning unit if it visited a known western Atlantic ICCAT spawning area (GOM, Bahamas or Florida Straits) for more than 7 days in winter or spring<sup>10,11,16–18</sup> and



Table 1 Electronic tags deployed in the western North Atlantic, 1996–2004

| Tag type (year)       | Releases | Tagged fish recaptures | Successful pop-ups | Recovery or reporting (%) | Mean length at tagging (cm CFL) |
|-----------------------|----------|------------------------|--------------------|---------------------------|---------------------------------|
| Archival* (1996–1999) | 280      | 77                     | n.a.               | 28                        | 199 ± 16                        |
| Archival† (2002–2004) | 219      | 9                      | n.a.               | 4.1                       | 199 ± 19                        |
| PAT (1997–2004)       | 273      | 12‡                    | 237                | 89                        | 211 ± 20                        |

n.a., not applicable.  
\* Archival tag models: NMT V1.1, V1.2 and WC Mk7.  
† Archival tag model: Lotek LTD2310.  
‡ Six recaptures of PAT-tagged fish occurred after the PAT tag had been released and are also recorded as successful pop-up reporting events. Six bluefin tuna recaptures before PAT tag release are not listed in successful pop-up events. One tag did not transmit and the PAT tag was recovered on a beach.



**Figure 1** Positions of Atlantic bluefin tuna electronically tagged at three western Atlantic locations (arrows) during 1996–2004. Circles represent locations based on deployment positions, light-based and SST-based geolocation estimates<sup>7–9</sup>, and PAT tag satellite endpoint positions. **a**, Fish classified as western breeders (10 archival tags, 26 PAT tags, 219 ± 27 cm CFL at release, median time at large 579 days). **b**, Fish classified as potential eastern breeders (23 archival tags, 3 PAT tags, 207 ± 17 cm CFL at release,

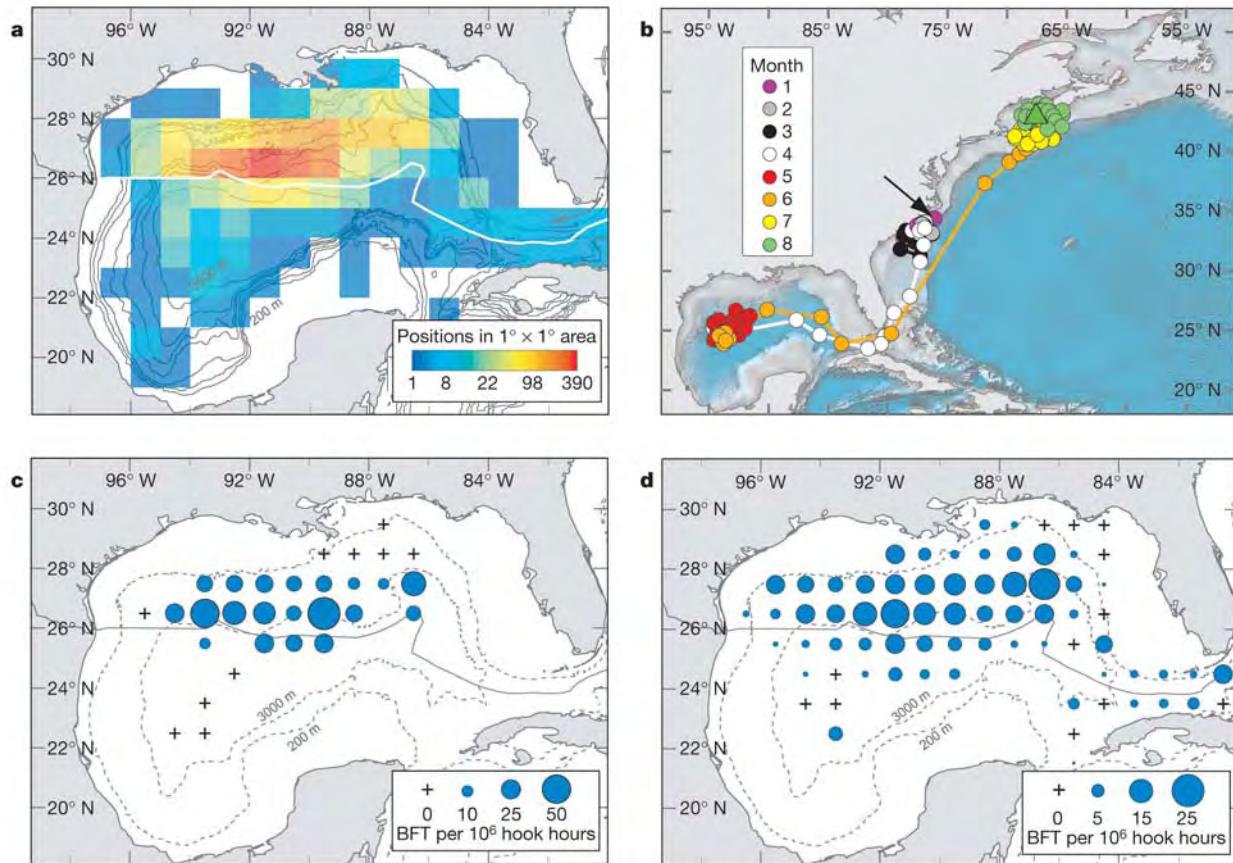
median time at large 926 days). **c**, Fish that did not visit a known ICCAT breeding ground (53 archival tags, 215 PAT tags, 202 ± 16 cm CFL at release, median time at large 141 days). **d**, Spatial overlap of western and eastern breeders identified in **a** and **b**. The dashed line in all panels indicates the current ICCAT management boundary (45° W meridian) and western breeding zone<sup>10,11</sup>. Triangles represent recapture locations of electronically tagged fish; the black triangle denotes  $n = 35$  recaptures.

occupied surface water temperatures of at least 24 °C, the SST reported for western spawning activity<sup>5,18,19</sup> ( $n = 36$ , Fig. 1a). Bluefin tuna that displayed transatlantic movements into the Mediterranean Sea and were recaptured in the spawning season (June to August<sup>10</sup>,  $n = 20$ ) or in the Straits of Gibraltar (May to August,  $n = 6$ ) were classified as potential eastern spawners (Fig. 1b). Bluefin tuna that remained in the North Atlantic throughout the track duration without visiting the known ICCAT spawning areas were classified as neutral ( $n = 268$ , Fig. 1c). We compared the distributions of the 62 bluefin tuna identified as potential western or eastern spawners and calculated a spatial overlap of the positional data sets of 47% in North Atlantic waters (Fig. 1d). These mixing zones were primarily in the western and central Atlantic. Importantly, no mixing occurred in the GOM and Mediterranean spawning areas.

Electronically tagged bluefin tuna were located in the GOM ( $n = 29$ , Figs 1a and 2a) from December to July. Electronically tagged fish were also located in the Bahamas ( $n = 6$ ) and northern Caribbean ( $n = 1$ ; Fig. 1a). The mean curved fork length (CFL) of electronically tagged bluefin that entered the GOM from the North Atlantic was  $241 \pm 28$  cm (about 11 years of age<sup>20</sup>). Location and diving data recorded on the tags<sup>5,6</sup> indicate that bluefin tuna enter the GOM along the continental slope through the Straits of Florida, diving to depths over 1,000 m (refs 5, 6), and move into the northern slope waters of the GOM. The mean SST recorded by

electronic tags on bluefin tuna in the GOM (Figs 1a and 2b), inclusive of transit and aggregation periods, was  $25.5 \pm 1.9$  °C. Electronic tag positions of bluefin tuna in the GOM, when combined with US pelagic longline observer and fisheries logbook bycatch data, identify areas of increased bluefin tuna occurrence ('hot spots') from 1992 to 2004 (Fig. 2). A majority of bluefin tuna locations (2,537 of 3,470; 73.1%) in the GOM from the three independent data sets were over the northern slope waters between the 200-m and 3,000-m contours (85° W to 95° W).

In the GOM slope waters, scientific longlining (live capture, tag and release) was conducted from pelagic longline vessels to tag giant bluefin tuna, and frequently resulted in bluefin tuna mortalities (Supplementary Table 1). This occurred even when short sets (less than 2 h soak time) and circle hooks (200 hooks or less) were used to reduce mortality. The mean size of bluefin tuna that died during capture on the longlines ( $237 \pm 17$  cm CFL,  $n = 16$ ) was similar to the mean size of bluefin tuna captured in the GOM by commercial fishers<sup>21</sup>. Histological examination indicates that ovaries from mortalities in the GOM in early April (1999 and 2000) were from mature fish in pre-spawn stages. Catches sampled in mid-April or May (2000 and 2001) revealed ovaries that were well vascularized with stages that included advanced yolke oocytes, oocytes with migrating nuclei, and post-vitellogenic oocytes exhibiting atresia. These ovarian stages were indicative of ripening, final maturation and post-ovulatory oocytes<sup>7,19,22</sup> and are consistent with previous



**Figure 2** Occurrence of Atlantic bluefin tuna on their western spawning ground in the Gulf of Mexico. **a**, Observed locations of Atlantic bluefin tuna in the GOM based on PAT tag satellite endpoint positions and geolocation estimates from electronic tags ( $n = 263$  positions, 1999–2004) and catch location statistics from pelagic longlines ( $n = 3,207$ , US scientific observer and US logbook data). **b**, Movements of an individual Atlantic bluefin tuna (03–251) showing a migration between the foraging grounds in the North Atlantic and the breeding grounds in the GOM. Colour denotes the month of each position. The bluefin tuna was released off North Carolina on 16 January 2004 (arrow, 268 cm CFL).

The tag detached from the fish on 27 August 2004 (green triangle). **c**, Distribution of Atlantic bluefin tuna CPUE in the GOM, based on the data from the US pelagic longline scientific observer program (1992–2004). **d**, Atlantic bluefin tuna CPUE, based on US pelagic longline logbook data (1992–2003). Only  $1^\circ \times 1^\circ$  areas with a total effort exceeding 50,000 and 500,000 hooks are shown in **c** and **d**, respectively. Areas exceeding this minimum effort without any bluefin tuna caught are denoted by black crosses. Solid white lines (**a**) and grey lines (**c** and **d**) indicate the US Exclusive Economic Zone.



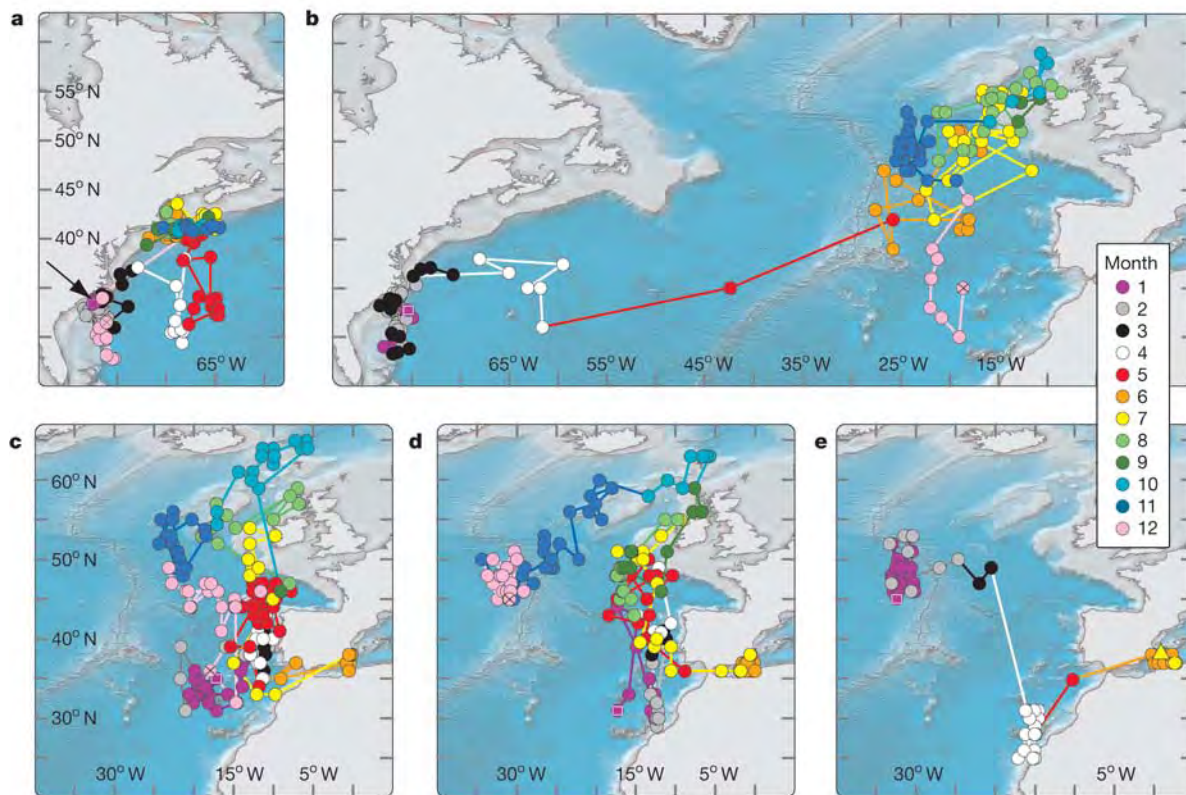
data indicating that spawning occurs in the GOM in April, May and June<sup>10,18</sup>. All histologically examined male testes ( $n = 6$ ) contained spermatozoa. Recent physiological studies indicate that Atlantic and Pacific bluefin tunas have an upper thermal tolerance for cardiac  $\text{Ca}^{2+}$  uptake that is crucial for heart function<sup>13,23</sup> (P. Castilho and B.A.B., unpublished data). The warm waters of the GOM are favourable for the development of the eggs and larval stages but may be physiologically stressful for giant tunas, which have high rates of heat production and large metabolic demands<sup>10</sup>. We propose that large bluefin tuna in spawning conditions might be susceptible to mortality on longlines in the GOM because of increased thermal and hypoxic stress induced by capture in warm surface waters.

After leaving the western spawning areas, the highest density of positions of bluefin tuna occurred in the waters overlying the North American continental shelf, slope and Gulf Stream waters, the South and mid-Atlantic Bight, the Gulf of Maine and the Nova Scotia Shelf (Fig. 1 and Supplementary Fig. 1). Another region of retention occurred in the central North Atlantic in the vicinity of 40° W, east of the Flemish Cap (Fig. 1). In this area, putative western spawners become vulnerable to central Atlantic fisheries of the eastern management unit.

Bluefin tuna ( $n = 26$ ) that had been electronically tagged in the western Atlantic showed transatlantic migrations to the Mediterranean Sea. These fish resided in the western Atlantic foraging grounds for 0.5 to 3 years before migrating to the Balearic Islands or the Tyrrhenian and/or Ionian seas (Figs 1b, 3 and 4, and Supplementary Fig. 2). These regions contain known spawning areas where mature females with hydrated oocytes, eggs and larvae

have been collected<sup>10,11,22</sup>. Bluefin tuna were recaptured in the Straits of Gibraltar ( $n = 5$ ) in May, potentially in transit to Mediterranean spawning areas, and in August ( $n = 1$ ), when tuna that have spawned might be re-entering the North Atlantic. The mean size at release in the western Atlantic of bluefin tuna that were recaptured in the Mediterranean Sea ( $n = 26$ ) was  $207 \pm 17$  cm CFL (about 8.6 years of age<sup>24</sup>). Western-tagged fish recaptured in the Mediterranean Sea seem to be returning to natal spawning areas. This hypothesis implies that a proportion of bluefin tuna electronically tagged in the western Atlantic are of eastern stock origin and are affecting western fisheries.

Spawning site fidelity to the Mediterranean Sea was evident for fish that were tagged in the western Atlantic and provided multi-year records (3.3–4.6 years). Bluefin tuna 603 (191 cm CFL, released on 17 January 1999) showed one year of western residency, a transatlantic crossing to the east Atlantic (2000), and three consecutive years (2001–2003) of summer movements into and out of the Mediterranean Sea, near the Balearic Islands (Fig. 3). Bluefin 705 (222 cm CFL, released on 11 February 1999) also showed spawning site fidelity to the western Mediterranean Sea during 2000–2003 (Supplementary Fig. 2). Bluefin 408 (203 cm CFL, released on 3 March 1997) spent three years foraging in the western Atlantic before a transatlantic migration into the Ionian Sea in 2000 (ref. 5). To date, only one bluefin tuna that went into the GOM (bluefin 512, 207 cm CFL, released on 17 January 1999) had a sufficiently long track to show western spawning site fidelity over two consecutive years<sup>5</sup> (S.L.H.T. and B.A.B., unpublished data). Multi-year records, although rare, reveal the complex ontogeny of movement patterns, which must be accounted for in stock management.



**Figure 3** Movements over 4.5 years of one individual Atlantic bluefin tuna (603) that was tagged in the western Atlantic in 1999 and demonstrated site fidelity to a known spawning area in the Mediterranean Sea (2001–2003). Each panel shows a year of the fish's track; colour denotes month of each position. Start and end points for each year are denoted by a square and cross-hatched circle, respectively. **a**, The bluefin tuna was released off North Carolina on 17 January 1999 (arrow, 191 cm CFL) and showed a year of western

residency. **b**, In 2000, the bluefin tuna showed transatlantic movement to the eastern Atlantic. **c–e**, Three consecutive years of movements from the eastern Atlantic into the Mediterranean Sea, to the vicinity of the Balearic Islands, during the breeding season: **c**, 2001; **d**, 2002; **e**, 2003. The fish was recaptured on 2 July 2003 (yellow triangle).

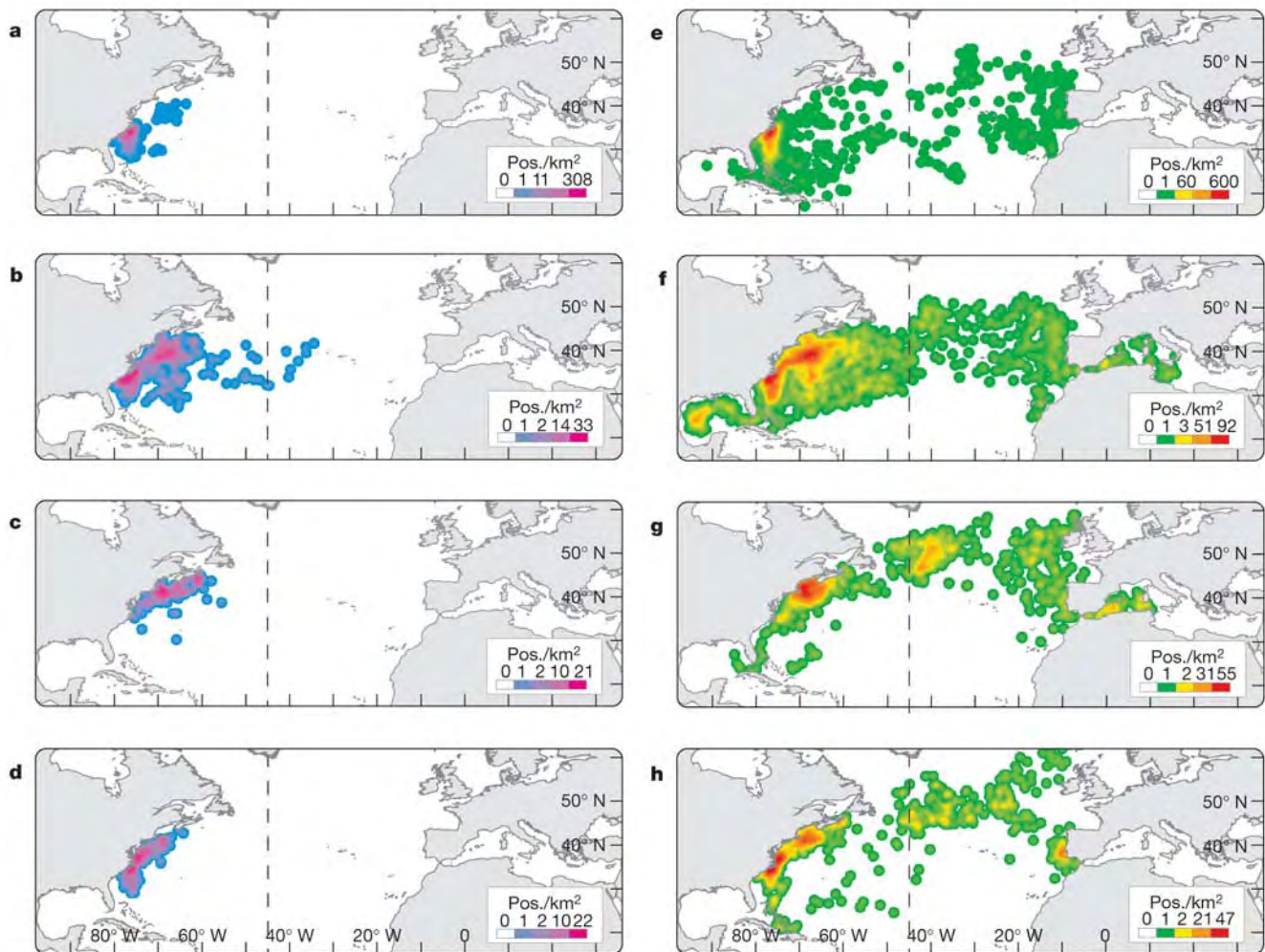


For fish that did not move into a known spawning ground ( $n = 268$ , Fig. 1c) the tracking durations were shorter (median duration 141 days) because of premature release of PAT tags or failure of sensors on early generation archival tags. We were therefore unable to discern whether, or when, these fish proceeded to breeding areas. Many fish were less than 200 cm CFL ( $n = 115$ ) and are, by length measurements, adolescent western fish. Some bluefin tuna with an unclassified breeding status ( $n = 35$ , at least 200 cm CFL) did experience SSTs of 24 °C or more in the waters of the South Atlantic or mid-Atlantic Bights and Gulf Stream. Ichthyoplankton surveys in the West Atlantic have captured bluefin tuna larvae off the Carolinas, although their presence was associated with advection from the Florida Straits and not from offshore spawning<sup>10,25</sup>. Examination of the ovaries of the bluefin tuna captured in the winter off North Carolina ( $n = 24$ , 195–227 cm CFL, January to April) has so far not revealed histological evidence of spawning adults in this region. However, these areas may represent extended ranges of western spawning areas<sup>4,5</sup> in late spring and early summer months, and require further study.

We examined the movements of electronically tagged bluefin tuna in relation to body size and season (Fig. 4). Bluefin tuna smaller than 200 cm CFL did not enter a known ICCAT spawning

area (Fig. 4a–d). Most of these fish remained west of 45° W throughout the year but displayed some range expansion in spring (Fig. 4a–d). Only bluefin tuna of larger body size (at least 200 cm CFL) occupied known spawning grounds from winter to early summer (western) and spring and summer (Mediterranean, Fig. 4e–h). This asynchrony in spawning is probably due to western spawning grounds acquiring optimal temperatures for bluefin spawning earlier than eastern spawning areas. In summer and autumn, fish of larger body size in both management units move into oceanic areas of high seasonal productivity at the northern extent of their range and along the continental shelves, while smaller bluefin remain primarily in areas along the North American shelf and slope waters (Fig. 4).

Archival tags, which have a large reward (US \$1,000) to increase recovery rates, demonstrate that 32 of the 86 recaptured bluefin tuna (37.2%) moved from the western to the eastern Atlantic management unit. Inclusion of PAT tags with shorter mean track durations yields a transfer rate west to east of 14.5% (48 of 330 fish). The probability of making a west-to-east transatlantic migration in all electronically tagged fish depends on the time at liberty, putative stock origin, and body size (Supplementary Table 2). In the first 6 months after tagging, bluefin tuna from both putative stocks had a



**Figure 4** Seasonal distribution by size of Atlantic bluefin tuna that were tagged in the western Atlantic and measured before release. **a–d**, Less than 200 cm CFL. **a**, Winter; **b**, spring; **c**, summer; **d**, autumn. **e–h**, Greater than or equal to 200 cm CFL. **e**, Winter; **f**, spring; **g**, summer; **h**, autumn. The dashed line in each panel indicates the current ICCAT management boundary (45° W meridian). High kernel densities<sup>29</sup> indicate seasonal hot spots where western-tagged Atlantic bluefin tuna spent the majority of time from 1996

to 2004. Only fish that were measured were used in this analysis. A western<sup>20</sup> or eastern<sup>24</sup> growth model was applied to obtain daily length after tagging. **a**,  $n = 101$ , mean size at release  $192 \pm 9$  cm CFL. **b**,  $n = 56$ ,  $192 \pm 6$  cm CFL. **c**,  $n = 22$ ,  $192 \pm 7$  cm CFL. **d**,  $n = 13$ ,  $187 \pm 8$  cm CFL. **e**,  $n = 162$ ,  $219 \pm 14$  cm CFL. **f**,  $n = 167$ ,  $220 \pm 13$  cm CFL. **g**,  $n = 97$ ,  $225 \pm 15$  cm CFL. **h**,  $n = 49$ ,  $227 \pm 15$  cm CFL. Pos., positions.

high probability of remaining in the western management unit (west,  $0.994 > P > 0.982$ ; east,  $0.933 > P > 0.900$ ; Supplementary Table 2, 95% confidence interval, 1,000 bootstrap samples). As time at liberty increases, the probability of remaining west of  $45^\circ\text{W}$  remains about the same for western fish but decreases rapidly for bluefin identified as eastern spawners (Supplementary Table 2). This result indicates that one component of the transatlantic migration is associated with fish of potential eastern origin moving back into the east Atlantic and Mediterranean Sea. A second component is associated with western breeding fish moving into eastern foraging grounds where encounters occur with eastern fishers (Fig. 1a).

The transatlantic movements observed in electronic tag data sets are corroborated by conventional tagging data, which demonstrate that 10% of tag recaptures from fish tagged and released in the South Atlantic Bight (1994–2000) occur in the eastern Atlantic and Mediterranean Sea<sup>5</sup>. Conventional tagging in the eastern Atlantic (1911–1990) indicated that 4.5% of recaptured juvenile bluefin tuna released in the eastern Atlantic were recaptured in the western Atlantic<sup>10</sup>. However, in these studies, no giant bluefin tuna conventionally tagged in the eastern Atlantic was recaptured in the western Atlantic<sup>10</sup>. Consistent with this result is the observation that no electronically tagged fish that moved into the Mediterranean Sea during spawning season has so far returned to the western Atlantic management unit. The conventional and electronic tagging data indicate that some juvenile fish tagged in the eastern Atlantic swim to the western Atlantic, where they remain for several years (Fig. 3 and Supplementary Fig. 2) before returning to Mediterranean spawning areas. We hypothesize that once an eastern spawned bluefin tuna returns from the North Atlantic to the Mediterranean it is less likely to forage along the North American coast. Fish identified as western spawners can move to the eastern Atlantic and back, crossing the  $45^\circ\text{W}$  meridian several times over the course of one or more years. The overlap areas identified in the central and eastern Atlantic seem to be foraging areas for these western spawners.

Five conclusions with management implications are apparent. First, our results support the existence of two North Atlantic bluefin tuna stocks, with discrete spawning areas primarily in the GOM and the Mediterranean Sea. Second, the two stocks overlap on North Atlantic foraging grounds as adolescents and adults, but there is no evidence for movement between the two major spawning areas in the GOM and the Mediterranean Sea. Third, fish identified as western or eastern spawners are subject to fishing pressures within their designated management unit during the spawning season. Fourth, the northern slope waters of the GOM are a critical habitat for bluefin tuna during the spawning season, and these fish could be protected with time-area closures to reduce the incidental catch of giant bluefin tuna by pelagic longline fisheries operating in the GOM. Fifth, transatlantic movements of western tagged fish have two components, one associated with tuna of eastern origin moving back to the Mediterranean spawning grounds, and another with western origin fish moving into eastern Atlantic foraging grounds.

Collaborative studies that combine electronic tagging data, otolith microchemistry<sup>26</sup> and genetics<sup>27</sup> should provide a method for validating and quantifying the extent of mixing between the putative stocks. Significant questions remain, including the relationship of the two North Atlantic bluefin tuna stocks tagged in the western Atlantic to the recently identified genetically distinct stock in the eastern Mediterranean Sea<sup>27</sup>. Quantifying the extent of spawning in one location relative to another, establishing whether individual adult bluefin tuna spawn every year and determining the influence of physical and biological oceanographic conditions on movements are essential to improved management strategies. If the electronic tagging results are used to develop and validate new models<sup>28</sup> of population mixing in the context of the dynamic North Atlantic environment, ICCAT will have a better opportunity to

prevent a further decline in the Atlantic Ocean's remaining bluefin tuna.

**Note added in proof:** During production of the manuscript, two additional tags were recaptured in December 2004 in the central Atlantic: LTD 2310 archival tag 781 at  $46.49^\circ\text{N}$ ,  $39.97^\circ\text{W}$ , and LTD 2310 tag 744 at  $44.50^\circ\text{N}$ ,  $30.28^\circ\text{W}$ . □

## Methods

Implantable archival tags were surgically placed in Atlantic bluefin tuna from 1996 to 2004 as described previously<sup>2,3,5,6</sup> (Table 1). Five models of archival tags (Northwest Marine Technology v1.1 and v1.2, Wildlife Computers Mk7 versions 1 and 2, and the Lotek LTD 2310) were deployed. Specifications of the tag sensors are available at the manufacturers' websites. Fishers reported 86 archival tags with corresponding conventional external tags (Table 1), but failed to return 20 electronic tags, which were included only as deployment and recovery positions. From 1997 to 2004 (Table 1), four generations of PAT tags<sup>5,6,9</sup> (Wildlife Computers, hardware versions 1 and 2, with modifications) were placed externally on bluefin tuna in North Carolina ( $n = 213$ ), Massachusetts ( $n = 33$ ) and the GOM ( $n = 27$ ). Pressure, light intensity, ambient and internal temperature data were recorded every 60, 120 or 128 s by the implantable archival tags. All longitude estimates were derived from light-intensity data recovered from or transmitted by the electronic tags, using threshold or template techniques<sup>7–9</sup>. Light-level geolocation estimates were made with manufacturers' proprietary software on-board the tag (NMT and Lotek tags) or by post-processing the data (Mk7 and PAT tags, Geocontrol v3.02 and WC-GPE Suite software). The daily SSTs were obtained from the archival tag data by extracting the ambient temperatures within 1 m of the surface<sup>9</sup>.

Pop-up satellite archival tags collected data at intervals of 60–120 s, summarized data into 2–24-h bins, and transmitted summary data to Argos satellites (PAT software versions 1.06, 1.07, 1.08 for PAT 1.0 generation tags, 2.03 and 2.04 in 2001, 2.07a in 2002, 2.08e in 2003 and 3.01d in 2004; Wildlife Computers). SSTs and thermal profiles of the water column were obtained from the profiles of depth–temperature data transmitted by the PAT tags. These data consist of the minimum and maximum temperatures at the surface, maximum depth, and six intermediate depths, over each data summary interval. All electronic tag data were corrected for pressure drift and thermal inertia<sup>9</sup>.

The SST data were combined with the corresponding light-level longitude estimates to obtain latitude estimates<sup>9</sup>. The daily maximum diving depths recorded by the tags were used to filter the geolocation estimates so that the maximum diving depth did not exceed the known bathymetry (inclusive of error estimates) at the geolocation estimate for the corresponding day. The accuracy of the geolocation estimates was validated with double-tagging experiments and by comparing the last position estimated from our algorithm with the recapture or PAT-tag endpoint positions from bluefin tuna<sup>9</sup>. On bluefin tuna ( $n = 11$ ; comparisons with recapture positions), archival tags have root-mean-square (r.m.s.) errors of  $0.78^\circ$  and  $0.90^\circ$  for longitude and latitude estimates, respectively. For PAT tags ( $n = 49$ ), the r.m.s. errors in the longitude and latitude estimates were  $1.30$  and  $1.89^\circ$ , respectively<sup>9</sup>. After the geolocation estimates were made, we assigned each bluefin tuna to a spawning unit: western, eastern or neutral as described above.

All fish tagged on the decks of sport fishing vessels were measured (cm curved fork length, CFL). The daily lengths of fish identified as western or eastern spawners were then calculated from the western<sup>20</sup> or eastern<sup>24</sup> growth models, respectively. The western growth model was also used for fish that were not assigned to a breeding stock. All results in this study are reported as means  $\pm$  s.d. When length information is provided in the text, only fish that were measured during tagging are included.

We calculated the fixed kernel density of the positions by size class (less than 200 cm CFL and 200 cm CFL or more) and season, to make nonparametric estimates of the spatial distributions of the fish<sup>29</sup> (Fig. 4 and Supplementary Fig. 1). The search radius was fixed at  $1.25^\circ$  for all kernel density calculations because this was the mean of the geolocation errors when the data from archival and PAT tags were combined. The kernel densities were calculated with the ArcGIS 9.0 Spatial Analyst (ESRI Inc.). The seasons were delineated by the equinoxes and solstices.

The spatial overlap between the western and eastern breeders (Fig. 1) was obtained by determining the area in which both western and eastern breeders were present. We divided the study area into  $1.25^\circ \times 1.25^\circ$  cells and identified cells that contained geolocations from both western and eastern breeders. For both populations we calculated their 95% fixed kernel spatial distributions, with smoothing parameters estimated by least-squares cross-validation<sup>29</sup>. This was done with ArcView 3.3 (ESRI Inc.) and the Animal Movement Extension 2.0 (P. N. Hooze and B. Eichenlaub). The percentage spatial overlap of their spatial distributions was then calculated as a proportion of their spatial distributions<sup>30</sup>.

The Atlantic bluefin tuna catch per unit effort (CPUE) in the GOM (Fig. 2) was calculated from data collected by the US pelagic longline scientific observer program (1992–2004) and the US pelagic longline logbook program (1992–2003). Both data sets were obtained from the US National Marine Fisheries Service. The CPUE for each  $1^\circ \times 1^\circ$  area was calculated if the effort in each area exceeded 50,000 and 500,000 hook hours for the observer and logbook data set, respectively. The yellowfin tuna CPUE for both data sets were also calculated for comparison (Supplementary Fig. 3).

Received 6 October 2004; accepted 17 February 2005; doi:10.1038/nature03463.

1. Metcalfe, J. D. & Arnold, G. P. Tracking fish with electronic tags. *Nature* **12**, 665–666 (1997).
2. Block, B. A., Dewar, H., Farwell, C. & Prince, E. A new satellite technology for tracking the movements of Atlantic bluefin tuna. *Proc. Natl Acad. Sci. USA* **95**, 9384–9389 (1998).
3. Block, B. A. *et al.* Archival tagging of Atlantic bluefin tuna (*Thunnus thynnus thynnus*). *Mar. Tech. Soc. J.* **32**, 37–46 (1998).
4. Luttcavage, M. E., Brill, R. W., Skomal, G. B., Chase, B. C. & Howey, P. W. Results of pop-up satellite



tagging on spawning size class fish in the Gulf of Maine: Do North Atlantic bluefin tuna spawn in the mid-Atlantic? *Can. J. Fish. Aquat. Sci.* **56**, 173–177 (1999).

5. Block, B. A. *et al.* Migratory movements, depth preferences, and thermal biology of Atlantic bluefin tuna. *Science* **293**, 1310–1314 (2001).
6. Stokesbury, M. J. W., Teo, S. L. H., Seitz, A., O'Dor, R. K. & Block, B. A. Movement of Atlantic bluefin tuna (*Thunnus thynnus*) as determined by satellite tagging experiments initiated off New England. *Can. J. Fish. Aquat. Sci.* **61**, 1976–1987 (2004).
7. Hill, R. D. & Braun, M. J. in *Electronic Tagging and Tracking in Marine Fisheries* (eds Sibert, J. R. & Nielsen, J. L.) 315–330 (Kluwer, Boston, Massachusetts, 2001).
8. Ekstrom, P. A. An advance in geolocation by light. *Mem. Natl Inst. Polar Res.* **58**, 210–226 (2004).
9. Teo, S. L. H. *et al.* Validation of geolocation estimates based on light level and sea surface temperature from electronic tags. *Mar. Ecol. Prog. Ser.* **283**, 81–98 (2004).
10. National Research Council, *An Assessment of Atlantic Bluefin Tuna* (National Academy Press, Washington DC, 1994).
11. ICCAT. *Report of the Standing Committee on Research and Statistics 2002–2003* (ICCAT, Madrid, 2003).
12. Carey, F. G. & Lawson, K. D. Temperature regulation in free-swimming bluefin tuna. *Comp. Biochem. Phys. A* **44**, 375–392 (1973).
13. Blank, J. M. *et al.* In situ cardiac performance of Pacific bluefin tuna hearts in response to acute temperature change. *J. Exp. Biol.* **207**, 881–890 (2004).
14. Myers, R. A. & Worm, B. Rapid worldwide depletion of predatory fish communities. *Nature* **423**, 280–283 (2003).
15. Fromentin, J. The East Atlantic and Mediterranean bluefin tuna stock management: uncertainties and alternatives. *Sci. Mar.* **67**, 51–62 (2003).
16. Rivas, L. R. A preliminary report on the spawning of the western north Atlantic bluefin tuna (*Thunnus thynnus*) in the Straits of Florida. *Bull. Mar. Sci. Gulf Carib.* **4**, 302–321 (1954).
17. Baglin, R. E. Reproductive biology of western Atlantic bluefin tuna. *Fish. Bull.* **80**, 121–134 (1982).
18. Mather, F. J., Mason, J. M. & Jones, A. C. *Historical Document: Life History and Fisheries of Atlantic Bluefin Tuna* (NOAA Tech. Memo. 370, NOAA, Miami, 1995).
19. Schaefer, K. M. in *Tuna: Physiology, Ecology, and Evolution* (eds Block, B. A. & Stevens, E. D.) 225–270 (Academic, San Diego, 2001).
20. Turner, S. C. & Restrepo, V. R. A review of the growth rate of west Atlantic bluefin tuna, *Thunnus thynnus*, estimated from marked and recaptured fish. *ICCAT Coll. Vol. Sci. Pap.* **42**, 170–172 (1994).
21. Nemerson, D., Berkeley, S. & Safina, C. Spawning site fidelity in Atlantic bluefin tuna, *Thunnus thynnus*: The use of size–frequency analysis to test for the presence of migrant east Atlantic bluefin tuna on Gulf of Mexico spawning grounds. *Fish. Bull.* **98**, 118–126 (2000).
22. Medina, A., Abascal, F. J., Megina, C. & Garcia, A. Stereological assessment of the reproductive status of female Atlantic northern bluefin tuna during migration to Mediterranean spawning ground through the Strait of Gibraltar. *J. Fish Biol.* **60**, 203–217 (2002).
23. Landeira-Fernandez, A. M., Morrisette, J. M., Blank, J. M. & Block, B. A. Temperature dependence of the Ca<sup>2+</sup> ATPase (SERCA2) in the ventricles of tuna and mackerel. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **286**, R398–R404 (2004).
24. Cort, J. L. Age and growth of the bluefin tuna, *Thunnus thynnus* (L.) of the northeast Atlantic. *ICCAT Coll. Vol. Sci. Pap.* **35**, 213–230 (1991).
25. McGowan, M. F. & Richards, W. J. Bluefin tuna, *Thunnus thynnus*, larvae in the Gulf Stream off the Southeastern United States: satellite and shipboard observations of their environment. *Fish. Bull.* **87**, 615–631 (1989).
26. Rooker, J. R., Secor, D. H., Zdanowicz, V. S., De Metrio, G. & Relini, L. O. Identification of Atlantic bluefin tuna (*Thunnus thynnus*) stocks from putative nurseries using otolith chemistry. *Fish. Oceanogr.* **12**, 75–84 (2003).
27. Carlsson, J. *et al.* Microsatellite and mitochondrial DNA analyses of Atlantic bluefin tuna (*Thunnus thynnus thynnus*) population structure in the Mediterranean Sea. *Mol. Ecol.* **13**, 3345–3356 (2004).
28. Apostolaki, P., Babcock, E. & McAllister, M. Further investigation of the effects of stock mixing on estimates of the size of North Atlantic bluefin tuna population using the six-area population dynamics model presented in SCRS/2002/088. *ICCAT Coll. Vol. Sci. Pap.* **56**, 1121–1133 (2004).
29. Worton, B. J. Kernel methods for estimating the utilization distribution in home-range studies. *Ecology* **70**, 164–168 (1989).
30. Atwood, T. C. & Weeks, H. P. Spatial home-range overlap and temporal interaction in eastern coyotes: the influence of pair types and fragmentation. *Can. J. Zool.* **81**, 1589–1597 (2003).

**Supplementary Information** accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank R. Rinaldo, E. Prince, A. Seitz, T. Sippel, R. Schallert, N. Tulloch, G. Rosenwaks, S. Beemer, G. Shillinger, C. Perle, S. Vermillion, J. Bonaventura, D. Barber, M. Orbach, J. Jenkins, G. Stuve, P. Wright, D. Britt, B. Eakes, C. Perry, D. Brower, W. Whippen, R. Whorley, R. Jansenius, G. Sharp, R. Hill, T. Lindstrom, P. Ekstrom, P. Manuel, R. Ruais and S. Loga. We are indebted to the late Richard Novak for his contributions and sacrifice on behalf of the Tag-A-Giant programme. We thank the National Marine Fisheries Service (NMFS) for providing access to the GOM scientific observer data and US pelagic longline logbook data. The Tag-A-Giant programme was supported by grants and donations from the Packard, Pew, MacArthur, Disney, Marine Ventures, Gordon and Betty Moore, and Monterey Bay Aquarium Foundations. This research was supported in part by the NOAA NMFS, the NSF and the National Fish and Wildlife Federation. We acknowledge the extensive cooperation of the commercial and recreational captains and crews of fishing vessels in North Carolina, New England, Nova Scotia, Louisiana and Texas. We thank N. Miyabe of the National Research Institute of Far Seas Fisheries, ICCAT, G. Dinatale, G. DeMetrio, M. de la Serna and the EU COPEMED programme for return of electronic tags.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to B.A.B. ([bblock@stanford.edu](mailto:bblock@stanford.edu)).

## ..... Learned kin recognition cues in a social bird

Stuart P. Sharp<sup>1</sup>, Andrew McGowan<sup>2</sup>, Matthew J. Wood<sup>3</sup>  
& Ben J. Hatchwell<sup>1</sup>

<sup>1</sup>Evolution and Behaviour Group, Department of Animal and Plant Sciences,  
University of Sheffield, Western Bank, Sheffield S10 2TN, UK

<sup>2</sup>Marine Turtle Research Group, Centre for Ecology and Conservation, University  
of Exeter in Cornwall, Tremough Campus, Penryn TR10 9EZ, UK

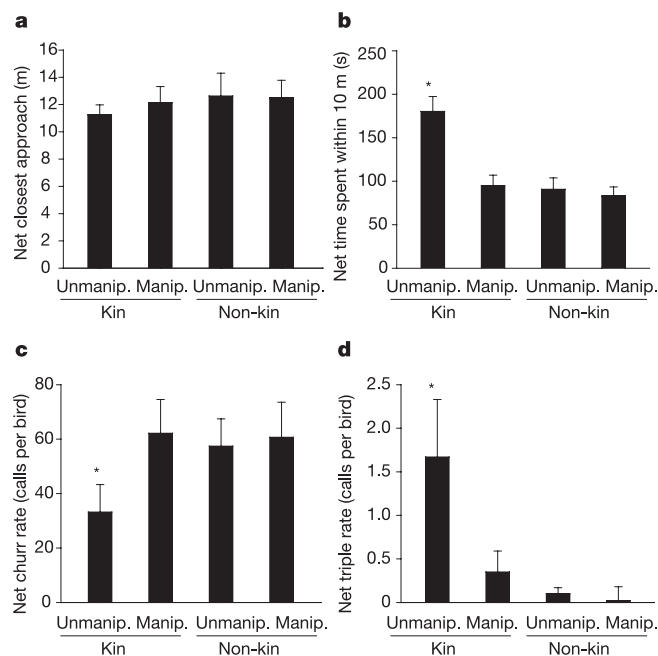
<sup>3</sup>The Edward Grey Institute of Field Ornithology, Department of Zoology,  
University of Oxford, South Parks Road, Oxford OX1 3PS, UK

In many cooperatively breeding birds, kin selection has an important role in the evolution and maintenance of social behaviour, and ‘helpers’ can maximize indirect fitness gains by preferentially allocating care to close relatives<sup>1–3</sup>. Although there is evidence for kin-biased helping behaviour in several species<sup>1,4,5</sup>, the mechanism of kin recognition underlying this behaviour is poorly understood<sup>2</sup>. Vocalizations are the most commonly used cues in avian recognition systems<sup>6,7</sup>, but the effectiveness of vocal signals as reliable recognition cues must depend on how they are acquired<sup>6–9</sup>. However, there have been no experimental studies of the development of vocal recognition cues in cooperative birds; indeed, the ontogeny of all bird vocalizations other than song is poorly known in any species<sup>10–12</sup>. Here, we show that cooperatively breeding long-tailed tits (*Aegithalos caudatus*) can discriminate between kin and non-kin according to the individual-specific characteristics of contact calls, and show experimentally that individuals learn these calls from provisioning adults during the nestling period. Finally, we show that the pattern of cooperative behaviour in this species is consistent with the use of recognition cues learned through association.

In long-tailed tits, all adults attempt to breed independently in pairs each year, but most nests fail due to depredation<sup>13,14</sup>. Failed breeders often re-nest, but later in the season may instead become helpers<sup>14</sup>; this switch from re-nesting to helping corresponds with a seasonal change in the potential fitness benefits of each strategy<sup>15</sup>. No significant direct fitness benefits of helping have been found, but helpers preferentially care for close relatives<sup>16</sup> and accrue indirect fitness benefits by increasing brood productivity<sup>14,15</sup>; this kin-selected benefit represents a substantial component of inclusive fitness and is the sole source of fitness for many individuals<sup>17</sup>. Thus, helping is beneficial to both helpers and recipients, and selection should favour kin recognition<sup>6,8</sup>. Kin-biased helping occurs in the absence of reliable spatial cues to kinship<sup>16</sup>, and a previous study suggested that long-tailed tits can discriminate between the vocalizations of close relatives and non-relatives<sup>18</sup>. Here, we describe an experiment that determines the characteristics of contact calls used in discrimination, and a second experiment that investigates the acquisition of these recognition cues.

Long-tailed tits have a limited vocal repertoire, with five call types and a very rarely used song<sup>13,19,20</sup>. The ‘churr’ call is a contact call given frequently by both sexes that is important for short-range communication; for example, during nest-building or aggressive interactions<sup>13,18–20</sup>. This call develops in the nest before fledging<sup>20</sup> and is highly stereotyped within individuals<sup>21</sup>, remaining unchanged throughout adulthood (S.P.S., unpublished data); multivariate analysis showed that maximum and minimum frequency are the two most individual-specific call parameters<sup>21</sup>. Using a playback experiment, we tested the ability of long-tailed tits to discriminate between the churr calls of kin and non-kin according to variation in these two parameters. We conducted playback trials with four treatments at the nests of focal birds using the following

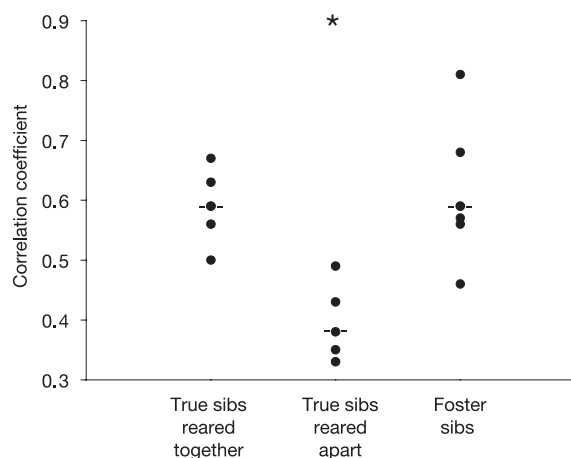




**Figure 1** Responses to playback trials ( $n = 8$ ) using the churr calls of kin and non-kin with maximum and minimum frequency unmanipulated (Unmanip.) and manipulated (Manip.). Net responses (error bars indicate mean  $\pm$  standard error) were calculated as the difference in response during playback and quiet periods. **a**, Closest approach to the speakers (Friedman test,  $s = 0.63$ ,  $P = 0.889$ ). **b**, Time spent within 10 m of the speakers ( $s = 14.85$ ,  $P = 0.002$ ). **c**, Churr rate ( $s = 12.15$ ,  $P = 0.007$ ). **d**, Triple rate ( $s = 15.93$ ,  $P = 0.001$ ). Tests remained significant after sequential Bonferroni correction. Asterisks indicate significant differences after treatment comparison tests<sup>28</sup>.

stimuli: (1) the churr calls of a close relative (coefficient of relatedness,  $r = 0.5$ ); (2) the churr calls from treatment 1 but with maximum and minimum frequency manipulated; (3) the churr calls of a non-relative ( $r < 0.125$ ); and (4) the churr calls from treatment 3 but with maximum and minimum frequency manipulated. The frequency parameters of manipulated calls remained within the range of natural variation observed in this species. The difference in each of four behavioural responses of focal birds during periods of playback and periods of quiet (that is, with no playback) was calculated to give four 'net' responses for each treatment. For three of the four responses measured there was a significant difference in net response during the four treatments (Fig. 1). In each case, the net response during playback of the unmanipulated churr calls of a close relative was significantly different from that during the other three treatments, between which there were no significant differences (Fig. 1). Long-tailed tits therefore responded differently to the manipulated and unmanipulated churr calls of kin, yet manipulation had no significant effect on the birds' responses to calls of non-kin. Thus, individuals were able to discriminate between the vocalizations of kin and non-kin based at least in part on variation in maximum and minimum frequency. This result does not imply that the churr call is the only kin recognition cue used by long-tailed tits: several different cues may be used, either in combination or separately according to context<sup>7,9</sup>. However, the results do show that the churr call alone is sufficient for successful discrimination.

The churr call may function as a vocal cue for kin recognition, but the reliability of such cues will depend in part on the nature of their development. The use of genetically determined cues may lead to recognition errors due to the effects of recombination, whereas cues derived from the environment are only reliable if acquired at a time when there is good evidence of kinship<sup>6-9</sup>. We conducted a cross-fostering experiment to investigate the relative contribution of

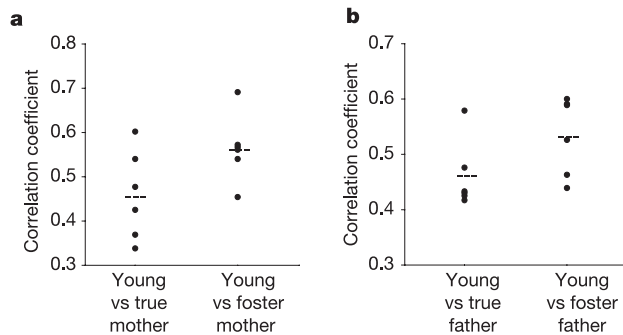


**Figure 2** Call similarity between different groups of siblings. Correlation coefficients (dashed horizontal bars indicate means) for pairwise comparisons of churr calls in each group were obtained using SPCC (Kruskal–Wallis test,  $\chi^2 = 9.752$ ,  $P = 0.008$ ). Asterisks indicate significant differences after treatment comparison tests<sup>28</sup>.

genetic and environmental influences on the development of the churr call. In a previous observational study, spectrographic cross-correlation (SPCC) revealed that the churr calls of siblings were more similar than those of non-siblings (mean  $\pm$  s.d. correlation coefficient for the calls of siblings =  $0.54 \pm 0.10$ ,  $n = 46$  pairs of siblings; for the calls of non-siblings =  $0.47 \pm 0.08$ ,  $n = 500$  pairs of non-siblings; S.P.S., unpublished data). The aims of the cross-fostering experiment were to compare the churr calls of foster siblings and true siblings, and the churr calls of fostered birds with those of their foster and biological parents. Nestlings from 24 partial broods were marked and swapped between synchronous nests of unrelated birds ( $r < 0.125$ ). The churr calls of recruits from cross-fostered broods were recorded in the following year when they had reached reproductive maturity and commenced breeding; these calls were then compared using SPCC. The churr calls of foster siblings were just as similar as those of true siblings reared together, whereas those of true siblings reared apart were significantly less similar (Fig. 2). Correlation coefficients for foster siblings and true siblings reared apart were comparable with those in the observational study for siblings and non-siblings, respectively. Furthermore, the churr calls of fostered individuals were significantly more similar to those of their foster parents than to those of their biological parents, whether comparisons were made with female (Fig. 3a) or male (Fig. 3b) parents. There must therefore be a significant learned component in the development of these calls.

Avian calls were traditionally thought to be genetically determined<sup>11,12</sup>, but our results support the more recent idea that learning can have an important role in call development<sup>10,12,22,23</sup>, just as it does in song development. Reliance on kin recognition cues that develop through learning may result in recognition errors (that is, the acceptance of non-kin as kin) if interactions with non-kin occur during cue development, or if social relationships are not effective predictors of genetic relatedness<sup>6,8</sup>. However, in long-tailed tits, extra-pair paternity and brood parasitism are rare<sup>24</sup> and, as in most cooperatively breeding birds, the association between offspring and their relatives is extended over a relatively long period<sup>25</sup>—the risk of making recognition errors is therefore reduced.

Long-tailed tit helpers exhibit a kin preference when controlling for spatial cues<sup>16</sup>, so if recognition is achieved through learning we would predict that the pattern of helping reflects associations during periods of call development. We tested this prediction by examining whether those helpers whose entire life history was documented became helpers at nests belonging to an individual with whom they had been associated during the nestling phase,



**Figure 3** Call similarity between recruits from cross-fostered broods and their true and foster parents. Correlation coefficients (dashed horizontal bars indicate means) for pairwise comparisons of churr calls were obtained using SPCC. **a**, Comparisons between the calls of recruits and their true and foster mothers (Wilcoxon's signed rank test,  $z = -2.201$ ,  $P = 0.028$ ). **b**, Comparisons between the calls of recruits and their true and foster fathers (Wilcoxon's signed rank test,  $z = -2.201$ ,  $P = 0.028$ ).

either as siblings or as a recipient or donor of care. In 57 out of 64 (89%) cases the helper assisted at least one breeder with whom it had been associated during the nestling phase, either as a sibling (38 out of 57, 67%), an offspring (13 out of 57, 23%), a parent (2 out of 57, 3%), a helper (1 out of 57, 1%), or as a recipient of helper care (3 out of 57, 5%). In 3 out of 64 (5%) cases, helpers assisted at nests belonging to a sibling of either a parent or a helper who fed it as a nestling. These instances suggest that kin recognition might occasionally be achieved indirectly through shared call characteristics, but we cannot exclude the possibility that there was some direct prior association that we had not observed. Finally, in just 4 out of 64 (6%) cases the helper and recipients were unrelated and we had no record of prior association, either direct or indirect, between the helper and the assisted breeders. In these few instances, it is possible that the helpers made recognition errors, as might be expected to occur in any recognition system<sup>8</sup>.

The cooperative associations of long-tailed tits are broadly consistent with a recognition mechanism of learning through direct association, as expected among cooperatively breeding birds<sup>2,16,18,23,26</sup>. Development of the churr call in the nest provides the opportunity to learn kin recognition cues from provisioning adults at a time when the presence of non-kin is unlikely. However, such cues would not function effectively as a kinship label for individuals who were not associated during the appropriate period of development because of the rapid diluting effect of learning from parents in an outbred population. Thus, this learning mechanism limits the pool of potential beneficiaries of kin-directed cooperation to the subset of kin within the population with whom the helper has had direct association. □

## Methods

### Field recordings and acoustic analysis

We studied a colour-ringed population of 63–90 pairs of long-tailed tits in Melton Wood, Doncaster, UK (53° 20' N, 1° 30' W) in 2001–03; relatedness was determined from pedigrees. Calls were recorded between February and June at a distance of <15 m using a Sennheiser MKH 416P48U (during 2001–02) or MKH 60P48 (2003) microphone. Recordings were made on one side of TDK type II SA cassettes using a Sony WM-D6C Walkman and digitized with 16-bit accuracy at a sampling rate of 22,050 Hz. Spectrograms were produced in Avisoft SASLab Pro (version 4.23b, 2003) using a 256-point fast Fourier transform length with a Hamming window function, 100% frame size and 75% window overlap.

### Playback experiment

In 2003, we identified focal nests ( $n = 8$ ) at which one parent, the 'target bird', had a close relative ( $r = 0.5$ ) alive whose churr call had been recorded; target birds were also randomly allocated a living non-relative ( $r < 0.125$ ). Each target bird was subjected to four playback treatments using churr calls: (1) kin; (2) manipulated kin; (3) non-kin; and (4) manipulated non-kin. For treatments 1 and 3, we randomly selected one churr call from recordings of the appropriate bird, then created a 1-min sequence of 36 randomly spaced

copies of the call (mean natural calling frequency in conspecific interactions at the nest =  $35.7 \pm 6.4$  calls per min,  $n = 50$ ). For treatments 2 and 4, sequences from treatments 1 and 3, respectively, underwent a frequency domain transformation in Avisoft, moving the entire spectrogram along the frequency axis by shifting a Fourier-transformed signal by a specified amount, then performing an inverse Fourier transformation. This procedure introduced no significant artefacts into the signal. The maximum frequency of churr calls ranges from 8.48 kHz to 10.47 kHz ( $n = 169$  individuals), the mid-point being 9.475 kHz. For calls with a maximum frequency above or below this mid-point, the spectrogram was decreased or increased by 1 kHz respectively; thus all calls remained within the normal frequency range. Sequences were then transferred to TDK endless cassettes.

We conducted trials by broadcasting calls through Sony SRS-58 speakers placed 10 m from focal nests containing nestlings. Treatments were run in different sequences at each nest, with two conducted on each of two consecutive days at the same times. Trials comprised 5 min of no playback followed by 5 min of playback. Target birds >20 m from the speakers were considered absent; if birds were absent throughout the quiet period the trial was restarted. An observer, who was unaware of which treatment was being conducted, stood 25–30 m from the nest and recorded the closest approach to the speaker and time spent <10 m from the speaker by the target bird. Calls could not always be assigned to individuals so the total numbers of churr calls and 'triple' calls (a long-range contact call<sup>13,18–20</sup>) were recorded and then divided by the number of birds present (two in most cases, but three for nests with a helper) to give the 'churr rate' and 'triple rate'.

### Cross-fostering experiment

In 2002, we marked partial broods (mean =  $4.50 \pm 0.66$  nestlings,  $n = 24$  broods; mean brood at hatching = 9.1 nestlings<sup>14</sup>) of 4–5-day-old nestlings and switched them between synchronous nests of unrelated birds. In 2003, the churr calls of philopatric recruits from cross-fostered nests were recorded and spectrograms produced of one randomly selected call per individual. Three categories were identified: (1) true siblings reared together; (2) true siblings reared apart; and (3) foster siblings. For dyads in each category, we compared their spectrograms by SPCC using Avisoft Correlator with a tolerated frequency deviation of 50 Hz and a high-pass filter of 1 kHz. Recordings of the churr calls of foster and biological parents were available for six recruits. We randomly selected one call from each foster parent and biological parent and compared their spectrograms with those of the corresponding fostered recruit using SPCC.

### Helper–breeder associations

Data were available from a long-term study (1994–2004) of a colour-ringed population of 18–68 pairs of long-tailed tits in the Rivelin Valley, Sheffield, UK (53° 12' N, 1° 34' W); relatedness was determined from pedigrees. On alternate days during the 16-day nestling period, provisioning behaviour was observed (usually for 1 h) and the identities and provisioning rates of all birds that fed nestlings recorded<sup>27</sup>. A complete history of associations as nestling and provisioning adult was determined for a total of 64 helpers who were first ringed as nestlings. A further five 'helpers' at four nests were excluded as they were observed to feed a brood on just one occasion despite extensive observations (mean =  $9.5 \pm 3.1$  h,  $n = 4$  nests; range 6–13 h); typical helpers have a provisioning rate of 5.0 feeds per hour to day-8 nestlings<sup>27</sup>.

Received 11 January; accepted 9 March 2005; doi:10.1038/nature03522.

- Dickinson, J. L. & Hatchwell, B. J. in *Ecology and Evolution of Cooperative Breeding in Birds* (eds Koenig, W. D. & Dickinson, J. L.) 48–66 (Cambridge Univ. Press, Cambridge, 2004).
- Komdeur, J. & Hatchwell, B. J. Kin recognition: function and mechanism in avian societies. *Trends Ecol. Evol.* **14**, 237–241 (1999).
- Stacey, P. B. & Koenig, W. D. (eds) *Cooperative Breeding in Birds: Long-term Studies of Ecology and Evolution* (Cambridge Univ. Press, Cambridge, 1990).
- Emlen, S. T. & Wrege, P. H. The role of kinship in helping decisions among white-fronted bee-eaters. *Behav. Ecol. Sociobiol.* **23**, 305–316 (1988).
- Komdeur, J. The effect of kinship on helping in the cooperative breeding Seychelles warbler (*Acrocephalus sechellensis*). *Proc. R. Soc. Lond. B* **256**, 47–52 (1994).
- Beecher, M. D. Kin recognition in birds. *Behav. Genet.* **18**, 465–482 (1988).
- Halpin, Z. T. in *Kin Recognition* (ed. Hepper, P. G.) 220–258 (Cambridge Univ. Press, Cambridge, 1991).
- Sherman, P. W., Reeve, H. K. & Pfennig, D. W. in *Behavioural Ecology: An Evolutionary Approach* 4th edn (eds Krebs, J. R. & Davies, N. B.) 69–96 (Blackwell, Oxford, 1997).
- Waldman, B. Mechanisms of kin recognition. *J. Theor. Biol.* **128**, 159–185 (1987).
- Hughes, M., Nowicki, S. & Lohr, B. Call learning in black-capped chickadees (*Parus atricapillus*): the role of experience in the development of 'chick-a-dee' calls. *Ethology* **104**, 232–249 (1998).
- Marler, P. & Mundinger, P. in *The Ontogeny of Vertebrate Behavior* (ed. Moltz, H.) 389–450 (Academic, New York, 1971).
- Baptista, L. F. in *Ecology and Evolution of Acoustic Communication in Birds* (eds Kroodsma, D. E. & Miller, E. H.) 39–60 (Cornell Univ. Press, Ithaca, 1996).
- Cramp, S. & Perrins, C. M. (eds) *The Birds of the Western Palearctic* Vol VII (Oxford Univ. Press, Oxford, 1993).
- Hatchwell, B. J. et al. Helpers increase long-term but not short-term productivity in cooperatively breeding long-tailed tits. *Behav. Ecol.* **15**, 1–10 (2004).
- MacColl, A. D. C. & Hatchwell, B. J. Temporal variation in fitness payoffs promotes cooperative breeding in long-tailed tits *Aegithalos caudatus*. *Am. Nat.* **160**, 186–194 (2002).
- Russell, A. F. & Hatchwell, B. J. Experimental evidence for kin-biased helping in a cooperatively breeding vertebrate. *Proc. R. Soc. Lond. B* **268**, 2169–2174 (2001).
- MacColl, A. D. C. & Hatchwell, B. J. Determinants of lifetime fitness in a cooperative breeder, the long-tailed tit *Aegithalos caudatus*. *J. Anim. Ecol.* **73**, 1137–1148 (2004).
- Hatchwell, B. J., Ross, D. J., Fowle, M. K. & McGowan, A. Kin discrimination in cooperatively breeding long-tailed tits. *Proc. R. Soc. Lond. B* **268**, 885–890 (2001).

19. Gaston, A. J. The ecology and behaviour of the long-tailed tit. *Ibis (Lond.)* **115**, 330–351 (1973).
20. Sharp, S. P. *Kin Recognition in the Cooperative Breeding System of the Long-Tailed Tit, Aegithalos caudatus*. Thesis, Univ. Sheffield (2003).
21. Sharp, S. P. & Hatchwell, B. J. Individuality in the contact calls of cooperatively breeding long-tailed tits. *Behaviour* (in the press).
22. Brittan-Powell, E. F., Dooling, R. J. & Farabaugh, S. M. Vocal development in budgerigars (*Melopsittacus undulatus*): contact calls. *J. Comp. Psychol.* **111**, 226–241 (1997).
23. Price, J. J. Family- and sex-specific vocal traditions in a cooperatively breeding songbird. *Proc. R. Soc. Lond. B* **265**, 497–502 (1998).
24. Hatchwell, B. J., Ross, D. J., Chaline, N., Fowle, M. K. & Burke, T. Parentage in the cooperative breeding system of long-tailed tits, *Aegithalos caudatus*. *Anim. Behav.* **64**, 55–63 (2002).
25. Hatchwell, B. J., Anderson, C., Ross, D. J., Fowle, M. K. & Blackwell, P. G. Social organization of cooperatively breeding long-tailed tits: kinship and spatial dynamics. *J. Anim. Ecol.* **70**, 820–830 (2001).
26. Komdeur, J., Richardson, D. S. & Burke, T. Experimental evidence that kin discrimination in the Seychelles warbler is based on association and not on genetic relatedness. *Proc. R. Soc. Lond. B* **271**, 963–969 (2004).
27. MacColl, A. D. C. & Hatchwell, B. J. Sharing of caring: nestling provisioning behaviour of long-tailed tits, *Aegithalos caudatus*, parents and helpers. *Anim. Behav.* **66**, 955–964 (2003).
28. Siegel, S. & Castellan, N. J. *Nonparametric Statistics for the Behavioral Sciences* (McGraw-Hill, Boston, 1988).

**Acknowledgements** We thank T.R. Birkhead, N.B. Davies, J. Slate and A.P. Beckerman for their advice; A.D.C. MacColl, D.J. Ross, M.K. Fowle and A.F. Russell for their assistance with fieldwork; and Doncaster and Sheffield City Councils, Yorkshire Water and Hallam Golf Club for allowing us to watch birds on their land. This research was supported by NERC and the University of Sheffield.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to S.P.S. (s.sharp@sheffield.ac.uk).

## A synthetic multicellular system for programmed pattern formation

Subhayu Basu<sup>1</sup>, Yoram Gerchman<sup>1</sup>, Cynthia H. Collins<sup>3</sup>,  
Frances H. Arnold<sup>3</sup> & Ron Weiss<sup>1,2</sup>

<sup>1</sup>Department of Electrical Engineering and <sup>2</sup>Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA

<sup>3</sup>Division of Chemistry and Chemical Engineering, California Institute of Technology 210-41, Pasadena, California 91125, USA

Pattern formation is a hallmark of coordinated cell behaviour in both single and multicellular organisms<sup>1–3</sup>. It typically involves cell–cell communication and intracellular signal processing. Here we show a synthetic multicellular system in which genetically engineered ‘receiver’ cells are programmed to form ring-like patterns of differentiation based on chemical gradients of an acyl-homoserine lactone (AHL) signal that is synthesized by ‘sender’ cells. In receiver cells, ‘band-detect’ gene networks respond to user-defined ranges of AHL concentrations. By fusing different fluorescent proteins as outputs of network variants, an initially undifferentiated ‘lawn’ of receivers is engineered to form a bullseye pattern around a sender colony. Other patterns, such as ellipses and clovers, are achieved by placing senders in different configurations. Experimental and theoretical analyses reveal which kinetic parameters most significantly affect ring development over time. Construction and study of such synthetic multicellular systems can improve our quantitative understanding of naturally occurring developmental processes and may foster applications in tissue engineering, biomaterial fabrication and biosensing.

Figure 1a depicts the design of the synthetic bacterial multicellular system, showing how only receivers at intermediate distances from senders express the output protein. Cell–cell communication from the senders is initiated by expression of the

LuxI enzyme<sup>4,5</sup> (Fig. 1b). LuxI catalyses the synthesis of AHL, which diffuses through the cell membrane and forms a chemical gradient around the senders. AHL diffuses into nearby receiver cells and is bound by LuxR, an AHL-dependent transcriptional regulator, which activates the expression of lambda repressor (CI) and Lac repressor (LacI<sub>M1</sub>, a product of a codon-modified *lacI*). Receiver cells in close proximity to the senders receive high concentrations of AHL, resulting in high cytoplasmic levels of CI and LacI<sub>M1</sub> and repression of the green fluorescent protein (GFP). Receivers that are far from the senders have low AHL concentrations, and accordingly LacI<sub>M1</sub> and CI are expressed only at basal levels. This enables the expression of a wild-type LacI, again resulting in GFP repression. At intermediate distances from the senders, intermediate AHL concentrations result in moderate levels of CI and LacI<sub>M1</sub>. However, because the repression efficiency of CI is significantly higher than that of LacI<sub>M1</sub>, CI effectively shuts off LacI expression while the LacI<sub>M1</sub> concentration is below the threshold required to repress GFP production. This difference between the CI and LacI<sub>M1</sub> repression efficiencies, in combination with a feed-forward loop<sup>6</sup> that begins with LuxR and culminates in GFP, affords the circuit the desired non-monotonic response to AHL dosages.

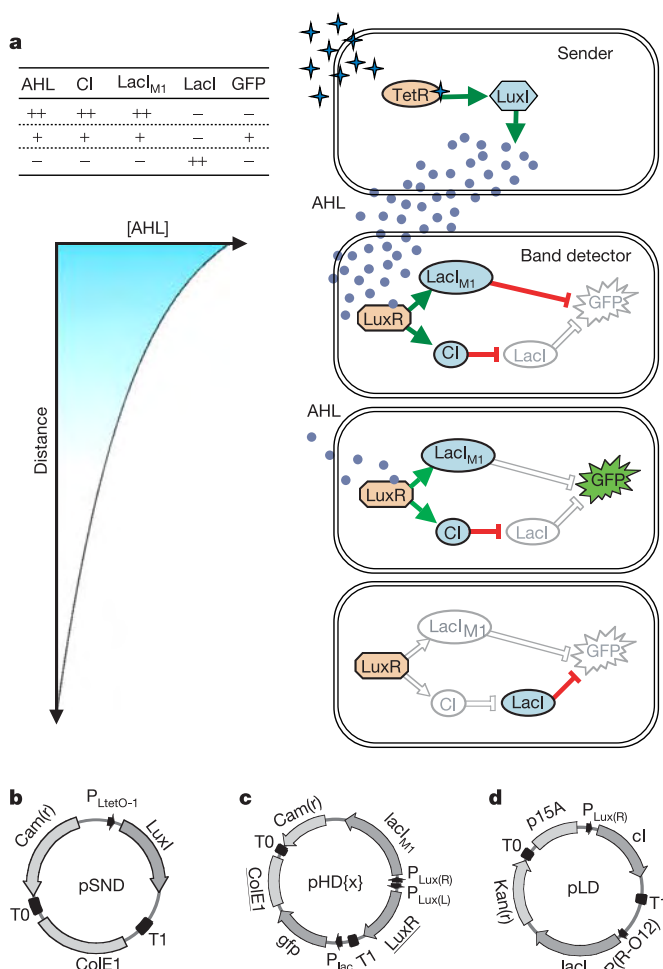
Guided by a mathematical model, the band-detect behaviour was engineered by combining a high-detect component (pHD plasmid; Fig. 1c) with a low-detect component (pLD plasmid; Fig. 1d) as described below. The high-detect component determines the AHL threshold above which GFP expression is muted. We engineered three high-detect strains (HD1, HD2 and HD3), each harbouring a variant of the high-detect plasmid (pHD{x}; Fig. 1c). The HD1 strain contains a hypersensitive LuxR mutant<sup>7</sup>, HD2 incorporates the wild-type LuxR, and HD3 cells express LuxR from a reduced-copy-number plasmid. In agreement with model predictions (Fig. 2a), the liquid-phase dosage responses of these three HD strains showed inverse correlations to AHL concentrations with different sensitivities (Fig. 2b). The low-detect component determines the lowest concentration of AHL that elicits GFP response. By combining the low-detect plasmid with each of the high-detect plasmid variants, we obtained three different band-detect strains named BD1, BD2 and BD3 accordingly. The BD strains showed a non-monotonic response to AHL with different thresholds (Fig. 2d), which correlated well with model predictions (Fig. 2c). Taken together, the responses of the three variants cover a wide range of biologically relevant AHL concentrations. Further analysis showing the effects of LacI and CI repression efficiencies on band-detect behaviour is included in the Supplementary Information.

Spatiotemporal simulations of a band-detect system predicted that by placing sender cells capable of AHL synthesis next to receiver cells, the above network could direct pattern formation on solid media. The model showed that given the appropriate kinetics for circuit elements, a distinct ring pattern would form in an initially undifferentiated ‘lawn’ of receiver cells around a group of sender cells (see Methods). Furthermore, a bullseye pattern could be achieved by mixing band-detect network variants such as BD1, BD2 and BD3. We tested these model predictions by plating on a Petri dish a mixture of BD3 cells and BD2-Red cells (similar to BD2 with *dsRed-Express* replacing *gfp*). A disk containing sender cells was placed in the middle of the dish (Fig. 3a), and the dish was incubated overnight. Microscope fluorescence images were subsequently captured. As seen in Fig. 3b, BD3 cells formed a green fluorescent ring near the senders, whereas BD2-Red cells formed a red fluorescent ring located further from the senders, creating a bullseye pattern. Similarly, when BD1 and BD2-Red cells were mixed and plated with a sender disk, an outer green fluorescent ring appeared around the red fluorescent ring (Fig. 3c). Although the relative positions of BD1, BD2 and BD3 cells were consistent in the two experiments, the diameters of the two BD2-Red rings were somewhat different (30 mm versus 22 mm). This can be attributed to variations in the AHL gradients due to differences in the growth rates and population



densities of senders and receivers, as well as subtle environmental differences between experiments such as nutrient conditions and agar densities.

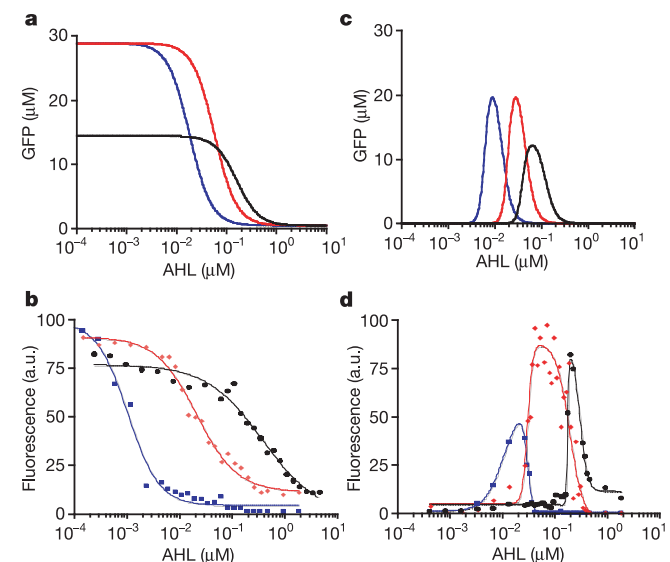
Understanding the dynamic behaviour of the system is crucial in predicting pattern formation. To study system dynamics, we measured the ring formation activity of BD2 cells over the course of 36 h. Fluorescence was recorded every 90 min for a 30 mm × 4 mm rectangular region protruding from the sender disk (Fig. 4a). For the first 15 h no fluorescence was observed, and then low levels of fluorescence emerged about 10 mm from the senders. Fluorescence values then increased significantly between 5 and 18 mm from the senders; these values stabilized after about 32 h, reaching a steady-state maximum at 10 mm (Fig. 4a). Over the duration of the experiment there was no observable shift in the position of high fluorescence.



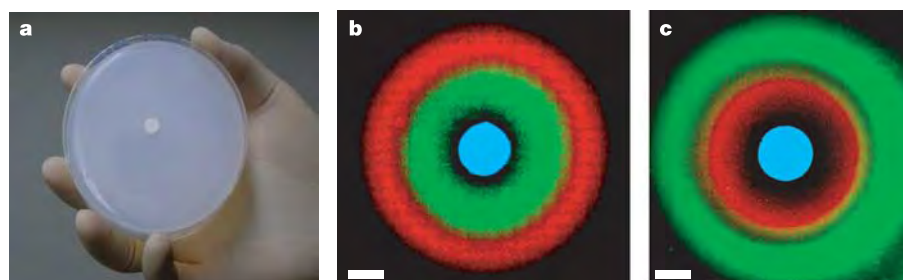
**Figure 1** The band-detect multicellular system programs *E. coli* receiver cells to fluoresce only at intermediate distances from sender cells. **a**, Circuit operation for a sender and three receivers exposed to high, medium or low AHL concentrations, showing the correlation of the various AHL and protein levels (top left), approximation of the AHL gradient as a function of the distance from the senders (bottom left) and the relevant protein activities in cells at different distances from the senders as mediated through transcriptional regulation (right; orange, constitutively expressed response proteins; blue/green, expression of regulated proteins; green and red arrows, transcriptional induction and repression respectively). High levels of LacI or LacI<sub>M1</sub> (indicated by ++) are required to repress GFP. **b**, Plasmid map for senders. **c, d**, The high-detect (**c**) and low-detect (**d**) plasmids that implement the band-detect operation. Three versions of the high-detect plasmid with different sensitivities to AHL were constructed (regions of mutation are underlined: pHD1, LuxR; pHD2, wild-type; pHD3, ColE1).

To gain a better understanding of circuit factors influencing ring formation dynamics, we performed and analysed 100 spatiotemporal simulations with different sets of kinetic rate constants. In all simulations, the senders' AHL synthesis rates were modified to ensure that fluorescence rings always formed at the same distance from the senders (as described in Methods). Figure 4b, c shows two simulations of the band-detect network with two different sets of kinetic rates. In comparison, the region of high fluorescence in Fig. 4c appears more quickly and stabilizes earlier, and its position shifts over a wider distance than in Fig. 4b. Statistical analysis of the entire set of simulations revealed that the rate constant for LacI decay had the strongest correlation with fluorescence response times and positional shift. These correlations are shown in Fig. 4d, e, where each point captures the response times (shift begin, shift end) and positional shift of one of the above simulations. The shift begin value is the time at which high fluorescence first appears, shift end is the time it takes the ring to form at the final position, and positional shift is the distance over which the ring expands until it stabilizes. The trend lines, computed as described in Methods, show a significant correlation between the rate constant for LacI decay and shift begin (Fig. 4d;  $R^2 = 0.91$ ,  $P < 0.0001$ ), shift end (Fig. 4d;  $R^2 = 0.71$ ,  $P < 0.0001$ ) and positional shift (Fig. 4e;  $R^2 = 0.76$ ,  $P < 0.0001$ ).

The stability of LacI affects how closely GFP expression in the receivers reports on the establishment of the AHL gradient from the senders. Before GFP expression can begin, AHL has to activate the production of CI, and LacI levels must subsequently decline. Fast rate constants for LacI decay will cause fluorescence to emerge quickly at a site far from the steady-state position, and then shift as the AHL gradient builds over time until it stabilizes. Circuits with intermediate LacI stability (decay rates between  $0.02 \text{ min}^{-1}$  and



**Figure 2** Simulated and experimental liquid-phase behaviour of high-detect and band-detect networks. **a, b**, Simulations (**a**) and experimental results (**b**) of the AHL dosage response for three high-detect network variants with wild-type LuxR (HD2, red), a hypersensitive LuxR (HD1, blue) and a reduced-copy-number plasmid (HD3, black). For the curve fits, the 95% confidence intervals have minimum/maximum values of 2.03/6.58, 2.47/4.14 and 2.78/5.12 for HD1, HD2 and HD3, respectively. **c, d**, Band detect simulations (**c**) and experimental results (**d**) of three networks consisting of the high-detect variants from above and the same low-detect component (BD1 (blue), BD2 (red) and BD3 (black) contain the high-detect components from HD1, HD2 and HD3, respectively). For the curve fits, the 95% confidence intervals have minimum/maximum values of 0.48/4.43, 7.36/18.55 and 1.1/8.23 for BD1, BD2 and BD3, respectively. a.u., arbitrary units.

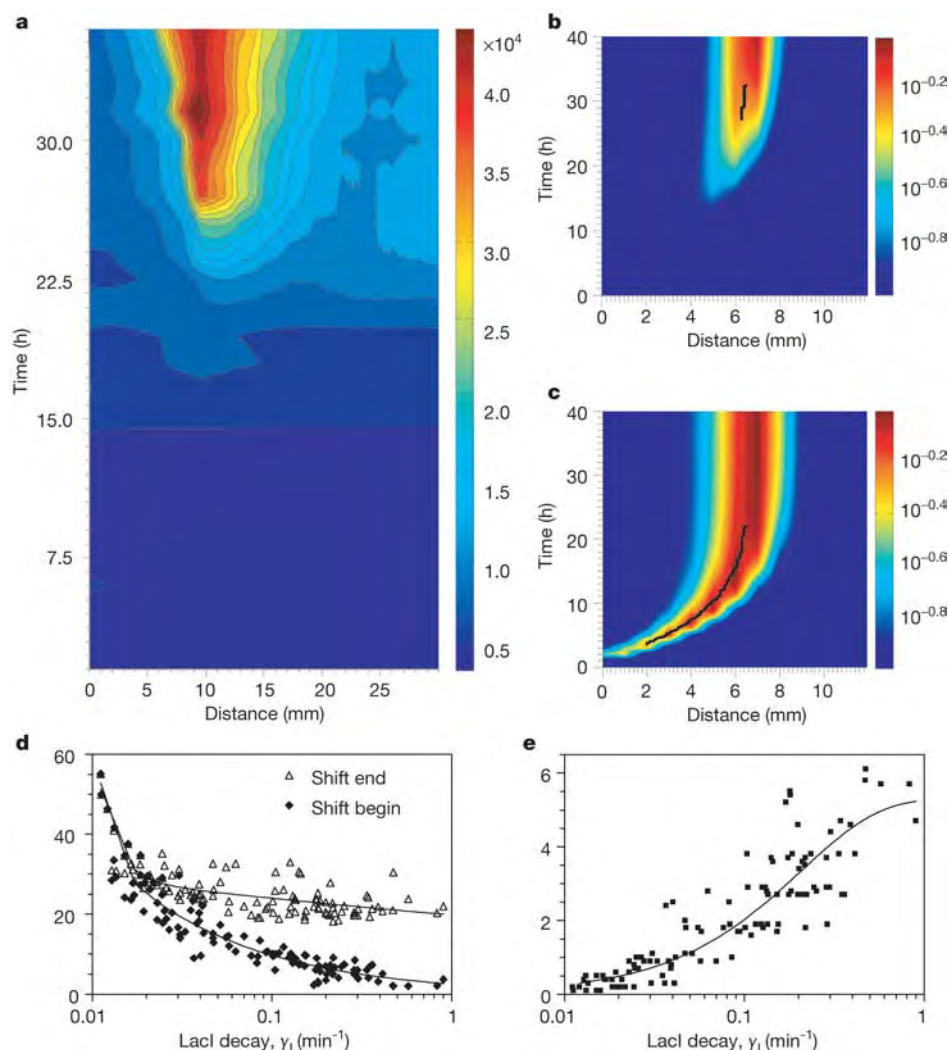


**Figure 3** Experimental solid-phase behaviour of band-detect networks. **a**, Picture of the Petri dish used in the BD2-Red/BD3 experiment showing the sender disk in the middle. **b**, Bullseye pattern as captured with a fluorescence microscope after incubation overnight with senders in the middle of an initially undifferentiated 'lawn' of BD2-Red and BD3 cells.

Surface maps depicting red and green fluorescence intensities are included in Supplementary Information. The senders in the middle are expressing CFP. **c**, Another bullseye pattern, this time with a mixture of BD1 and BD2-Red cells. Scale bar, 5 mm.

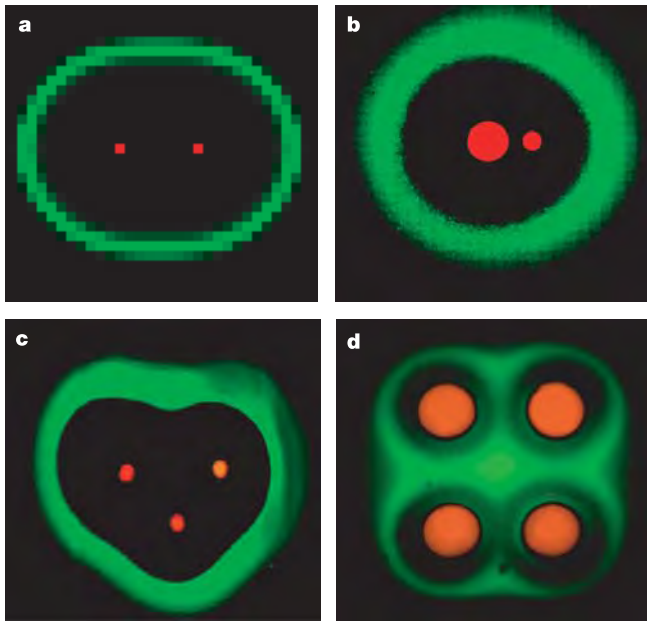
$0.04 \text{ min}^{-1}$ ) form rings at the final position only, but do not incur a significant time delay. Rate constants for LacI decay of less than  $0.02 \text{ min}^{-1}$  also show no positional shift but result in a very slow emergence of fluorescence long after the AHL gradient has stabilized. The effect of LacI stability is also demonstrated by

comparing Fig. 4b with Fig. 4c, where an important difference between these two kinetic parameter sets is a shorter half-life for LacI in Fig. 4b. We therefore postulated that the relatively long half-life of LacI (in our experimental system LacI degradation is due mainly to dilution by cell growth) accounts for the similarity



**Figure 4** Ring formation dynamics. **a**, Experimental results showing the time-evolution of fluorescence for band-detect cells as a function of the distance from the senders. **b**, **c**, Spatiotemporal simulations of two shifts with different sets of kinetic parameters that form the ring at the same distance. Maximal levels of fluorescence are indicated by red, while the black lines represent the spatiotemporal shift of the ring. The shift associated

with a fast decaying LacI (**c**) is larger than the shift resulting from a stable LacI (**b**). **d**, Regression analysis correlating the fluorescence response times with rate constants for LacI decay. Open triangles, shift end; filled diamonds, shift begin. **e**, Regression analysis correlating positional shift with rate constants for LacI decay.



**Figure 5** Formation of various patterns. **a**, Simulation of band-detect behaviour on solid media with two senders that results in the formation of an ellipse. **b–d**, Experimental results showing various GFP patterns formed based on the placement and initial concentrations of sender cells expressing DsRed-Express: **b**, ellipse, two sender disks; **c**, heart, three sender disks; and **d**, clover, four sender disks.

between the experimental observations in Fig. 4a and the results of the simulation in Fig. 4b. For the other rate constants, CI repression efficiency showed a weak correlation with the positional shift ( $R^2 = 0.05$ ,  $P = 0.0191$ ), whereas the rest of the rate constants from equations (1)–(4) did not show any correlation ( $R^2 < 0.001$ ,  $P > 0.75$ ).

The dynamics and final pattern of differentiation depend not only on the receiver network but also on the spatial arrangement of senders. We directed this multicellular system to form more elaborate patterns by placing multiple sender disks in different configurations. By changing initial conditions (for example the number of, distance between, and density of sender cells) one can create different AHL gradients and hence direct the formation of different intricate patterns (Fig. 5). In the future, more complex multicellular patterns with improved properties will be achieved by integrating bi-directional communication.

Natural systems are often a good source of inspiration for the forward design of synthetic circuits. For example, in *Drosophila melanogaster* blastoderm segmentation, inhibitory feedback loops between communicating elements help to demarcate domain boundaries<sup>8</sup>. By incorporating similar intercellular feedback loops, future artificial multicellular systems could create patterns with sharper edges. However, the feedback loops found in *D. melanogaster* are also responsible for ‘domain shifts’<sup>9</sup>. Although some patterns of gene expression are spatiotemporally stable (Kruppel), others shift over time (giant, knirps)<sup>9</sup>. Such a shift might be undesirable and even destructive in certain synthetic systems, for example tissue-engineering applications. Nevertheless, as shown by the statistical analysis above, this problem can be overcome by careful engineering of system components. The construction and analysis of artificial systems such as the band-detect network can also help to improve our understanding of biological design principles. These domain shifts in *D. melanogaster* are accompanied by a sharpening of the gene expression zones and probably represent an important fine tuning of the location and size of these zones. We suggest that such natural patterning systems have evolved finely

tuned protein decay rates that allow them to function with the precision required during development. It will be interesting to see whether the shift/no-shift property is indeed controlled by careful matching between intracellular protein decay dynamics and extracellular diffusion processes.

The work described here shows the design and construction of an artificial multicellular system capable of programmed pattern formation. We have shown how a community of cells can sense a chemical gradient to form three distinct regions. The system consists of simple parts that are arranged in different configurations to elicit the desired patterns. Theoretical and experimental analyses of system behaviour are facilitated by the fact that the parts are well characterized and can be fine-tuned. The integration of such systems into higher-level organisms and with different cell functions will have practical applications in three-dimensional tissue engineering, biosensing, and biomaterial fabrication. We see the construction of this and similar systems as a step towards creating artificial differentiation patterns on demand and contributing to a better understanding of natural developmental processes. □

## Methods

### Plasmids

Plasmids and their properties are listed in Supplementary Table S1. The sender plasmid pSND encodes LuxI controlled by  $P_{\text{LTet-O1}}$  promoter<sup>10</sup>. Band-detect cells contain two plasmids, low-detect plasmid (pLD) and high-detect plasmid (pHD[x]). pLD encodes a destabilized  $\lambda$  repressor<sup>11</sup> (CI) controlled by  $P_{\text{lux(R)}}$  (ref. 12) and a *lac* repressor (LacI) controlled by  $\lambda_{\text{P(R-O12)}}$  (ref. 13), and was constructed from pLTSUB-202 (ref. 14) and pRCV-3 (ref. 10). pHD2 was made from pRCV-102 (ref. 15) and pINV-4 (ref. 15), and contains LuxR,  $\text{LacI}_{\text{M1}}$  and GFP under the control of  $P_{\text{lux(L)}}$ ,  $P_{\text{lux(R)}}$  and  $P_{\text{lac}}$  respectively.  $\text{LacI}_{\text{M1}}$  is a codon-modified *lacI* designed to reduce the likelihood of recombination with the wild-type *lacI* on the pLD plasmid. pHD1 is a variant of pHD2 with a hypersensitive variant of LuxR, LuxR-G2F (ref. 7). pHD3 is a variant of pHD1 with an AT  $\rightarrow$  TA mutation in *ColE1 ori* positions 222–223 that approximately halves the plasmid copy number (data not shown). pHD2-Red was made from pHD2 by replacing GFP with DsRed-Express. All proteins except LuxR are preceded by a strong synthetic ribosome-binding site, RBSII<sup>13</sup>, whereas LuxR is preceded by its native RBS. CI and GFP are destabilized with an 11-amino-acid ssrA LVA tag<sup>16</sup>.

### Data acquisition and analysis

*Escherichia coli* strain DH5 $\alpha$  ( $\lambda^-$ ,  $\text{recA1}^-$ ) (CGSC strain 7855) was used for all experiments. Cells transformed with the appropriate plasmids were grown at 37 °C in M9 minimal medium (0.2% casamino acids, 200  $\mu\text{M}$  thiamine, 100  $\mu\text{M}$   $\text{CaCl}_2$ ), with antibiotics (50  $\mu\text{g ml}^{-1}$  kanamycin, 25  $\mu\text{g ml}^{-1}$  chloramphenicol) until exponential phase (optical density at 600 nm (OD 600) = 0.3). For the liquid-phase experiments, expression was induced by the addition of 3-oxohexanoyl homoserine-lactone (3OC<sub>6</sub>HSL; Sigma-Aldrich). Receiver cultures were analysed by fluorescence-activated cell sorting (FACS) with a Beckman Coulter Altra (488-nm argon excitation laser, 515–545-nm emission filter) and measurements were calibrated using Spherotech SPHERO Rainbow Calibration Particles (RCP-30-5A). Data points represent median FACS values. Figure 2c curves were fitted to a sigmoidal function,

$$y = y_0 + \left( \frac{\alpha}{1 + (\text{AHL}/\beta)^n} \right)$$

and Fig. 2d curves were fitted to a bell-shaped dose-response function,

$$y = \delta + \left( \frac{(\alpha_1 - \delta)}{1 + (\text{AHL}/\beta_1)^{n_1}} \right) + \left( \frac{(\alpha_2 - \delta)}{1 + (\text{AHL}/\beta_2)^{-n_2}} \right)$$

using GraphPad PRISM software. For the solid-phase experiment, time-lapse microscopy was conducted on a Zeiss Axiovert 200M microscope equipped with a 1344-pixel  $\times$  1024-pixel cooled ORCA-ER charge-coupled device camera (Hamamatsu). Separate cultures of band-detect cells were grown to an OD 600 of 0.3, washed with M9 salts, diluted into 2 ml of 0.7% agarose to achieve an OD 600 of 0.15, and spread evenly on top of an M9 1.5% agarose plate. Senders were grown in a similar manner and then concentrated 50-fold. Senders (20  $\mu\text{l}$ ) were pipetted onto a Whatman 3 paper disk and placed in the centre of the plate. Plates were incubated at 37 °C. Brightfield and fluorescence images were captured with a 2.5 $\times$  brightfield objective. False colouring was performed by capturing images with filters optimized for the different fluorescent proteins and then superimposing the images with the appropriate colours and assembling them into larger mosaics with custom software (CFP filters, 436/20 excitation and 470/30 emission; GFP, 470/40 and 525/50; DsRedExpress, 565/30 and 620/60).

### Model

A simple five-species model was used to model both the single-cell band-detect response to AHL and the spatiotemporal behaviour of a multicellular sender–receiver system (using the modelling tool described previously<sup>14</sup>). The models were based on ordinary differential equations with Hill functions that captured the activation and repression of protein synthesis. The intracellular species included GFP (G), LacI (L), CI (C), LuxR/AHL



complex (R), AHL (A) and a fixed concentration of LuxR. The following equations were used:

$$\frac{dG}{dt} = \frac{\alpha_G}{1 + (\beta_L/L)^{\eta_1}} - \gamma_G G \quad (1)$$

$$\frac{dL}{dt} = \frac{\alpha_{L1}}{1 + (C/\beta_C)^{\eta_2}} + \frac{\alpha_{L2} \cdot R^{\eta_3}}{(\theta_R)^{\eta_3} + R^{\eta_3}} - \gamma_L L \quad (2)$$

$$\frac{dC}{dt} = \frac{\alpha_C R^{\eta_3}}{(\theta_R)^{\eta_3} + R^{\eta_3}} - \gamma_C C \quad (3)$$

$$\frac{dR}{dt} = \rho_R [\text{LuxR}]^2 A^2 - \gamma_R R \quad (4)$$

$$\frac{dA_{x,y,z}}{dt} = \xi(A_{x-1,y,z} + A_{x+1,y,z} + A_{x,y-1,z} + A_{x,y+1,z} + A_{x,y,z-1} + A_{x,y,z+1} - 6A_{x,y,z}) - \gamma_A \quad (5)$$

The following parameters approximate the behaviour of BD2: protein synthesis rates ( $\alpha_G$ ,  $2 \mu\text{M min}^{-1}$ ;  $\alpha_{L1}$ ,  $1 \mu\text{M min}^{-1}$ ;  $\alpha_{L2}$ ,  $1 \mu\text{M min}^{-1}$ ;  $\alpha_C$ ,  $1 \mu\text{M min}^{-1}$ ), repression coefficients ( $\beta_L$ ,  $0.8 \mu\text{M}$ ;  $\beta_C$ ,  $0.008 \mu\text{M}$ ), protein decay ( $\gamma_G$  and  $\gamma_C$ ,  $0.0692 \text{ min}^{-1}$ ;  $\gamma_L$  and  $\gamma_R$ ,  $0.0231 \text{ min}^{-1}$ ), LuxR/AHL activation coefficient ( $\theta_R$ ,  $0.01 \mu\text{M}$ ), transcription factor cooperativity/multimerization ( $\eta_1$ , 2;  $\eta_2$ , 2;  $\eta_3$ , 1), LuxR/AHL dimerization ( $\rho_R$ ,  $0.5 \mu\text{M}^{-3} \text{ min}^{-1}$ ), AHL intercellular diffusion ( $\xi$ ,  $0.001 \text{ mm}^2 \text{ min}^{-1}$ ) and AHL decay ( $\gamma_A$ ,  $0.01 \text{ min}^{-1}$ ). LuxR concentration was set at  $0.5 \mu\text{M}$ . LuxR and AHL first bind to form a complex, and this complex then dimerizes to form an active transcription factor<sup>17</sup>; hence the quadratic terms in equation (4). The degradation rate of AHL ( $\gamma_A$ ), which determines the steepness of the chemical gradient, is affected by pH (ref. 18). The liquid-phase simulations did not include equation (5). Further, to simulate the high-detect component in the liquid phase, equation (3) and CI repression in equation (2) were also not used. For the HD1 and BD1 liquid simulations,  $\rho_R$  was increased tenfold from above, whereas for the HD3 and BD3 simulations the values of  $\alpha_G$ ,  $\alpha_{L2}$  and [LuxR] were halved.

For the statistical analysis of the shift, we first generated 2,000 sets of random kinetic parameters, in which the values for each parameter were uniformly distributed around those used above. Among these sets, about 30% yielded band-detect behaviour with a gain greater than three and fluorescence values above a predetermined threshold. For the first 100 sets we computed the AHL synthesis rate resulting in ring formation centred at 7 mm from the senders. Using each parameter set, we simulated system behaviour until GFP at all positions stabilized, and computed the shift as follows. The beginning of a shift is defined as the spatiotemporal coordinate  $\langle \text{position}_{\text{start}}, \text{time}_{\text{start}} \rangle$  that meets the following criteria: it is in the middle of the ring for that particular time, it is closest to the origin, and its fluorescence is greater than 50% of maximum overall steady-state fluorescence. The end of a shift is defined as the spatiotemporal coordinate  $\langle \text{position}_{\text{steady}}, \text{time}_{\text{steady}} \rangle$  that meets the same criteria as above and is located at the final steady-state position of the ring. The positional shift is then the difference between  $\text{position}_{\text{steady}}$  and  $\text{position}_{\text{start}}$ .

In Fig. 4b, c the following values were changed from above ( $\beta_L$ , 543/230 nM;  $\beta_C$ , 19/69 nM;  $\gamma_G$ , 0.116/0.496  $\text{min}^{-1}$ ;  $\gamma_C$ , 0.404/0.366  $\text{min}^{-1}$ ;  $\gamma_L$ , 0.018/0.902  $\text{min}^{-1}$ ;  $\gamma_R$ , 0.036/0.050  $\text{min}^{-1}$ ;  $\theta_R$ , 0.249/0.063  $\mu\text{M}$ ;  $\rho_R$ , 2.102/3.374  $\mu\text{M}^{-3} \text{ min}^{-1}$ ). GraphPad PRISM software was used to perform a regression analysis correlating each of the rate constants with the corresponding shift values. The response times were correlated by using power-series equations and the positional shift was correlated with a sigmoidal fit.

Received 3 December 2004; accepted 14 February 2005; doi:10.1038/nature03461.

- Golden, J. & Yoon, H. Heterocyst formation in *anabaena*. *Curr. Opin. Microbiol.* **1**, 623–629 (1998).
- Scherrer, R. & Shull, V. Structure, partial elemental composition, and size of *Thiopedia rosea* cells and platelets. *Can. J. Microbiol.* **32**, 607–610 (1986).
- Ben-Jacob, E. *et al.* Cooperative formation of chiral patterns during growth of bacterial colonies. *Phys. Rev. Lett.* **75**, 2899–2902 (1995).
- Fuqua, W. C., Winans, S. & Greenberg, E. P. Quorum sensing in bacteria: The luxR-luxI family of cell density-responsive transcriptional regulators. *J. Bacteriol.* **176**, 269–275 (1994).
- Bassler, B. L. How bacteria talk to each other: regulation of gene expression by quorum sensing. *Curr. Opin. Microbiol.* **2**, 582–587 (1999).
- Milo, R. *et al.* Network motifs: simple building blocks of complex networks. *Science* **298**, 824–827 (2002).
- Collins, C. H., Arnold, F. H. & Leadbetter, J. R. Directed evolution of *Vibrio fischeri* LuxR for increased sensitivity to a broad spectrum of acyl-homoserine lactones. *Mol. Microbiol.* **55**, 712–723 (2005).
- Jaeger, J. *et al.* Dynamical analysis of regulatory interactions in the gap gene system of *Drosophila melanogaster*. *Genetics* **167**, 1721–1737 (2004).
- Jaeger, J. *et al.* Dynamic control of positional information in the early *Drosophila* embryo. *Nature* **430**, 368–371 (2004).
- Weiss, R. & Knight, T. F. Jr in *DNA6: Sixth International Workshop on DNA-Based Computers, DNA2000* (eds Condon, A. & Rozenberg, G.) 1–16 (Springer, Leiden, The Netherlands, 2000).
- Ptashe, M. A. *Genetic Switch: Phage Lambda and Higher Organisms*, 2nd edn (Cell Press and Blackwell Scientific Publications, Cambridge, Massachusetts, 1986).
- Egland, K. A. & Greenberg, E. P. Quorum sensing in *Vibrio fischeri*: elements of the luxI promoter. *Mol. Microbiol.* **31**, 1197–1204 (1999).
- Weiss, R. & Basu, S. in *NSC-1: The First Workshop of Non Silicon Computing* (<http://www-2.cs.cmu.edu/~phoenix/nsc1/paper/3-2.pdf>) (2002).
- Basu, S., Mehreja, R., Thiberge, S., Chen, M. & Weiss, R. Spatiotemporal control of gene expression with pulse-generating networks. *Proc. Natl Acad. Sci. USA* **101**, 6355–6360 (2004).
- Weiss, R. *Cellular Computation and Communication Using Engineered Genetic Regulatory Networks*. Ph.D. thesis, Massachusetts Inst. Technology (2001).

- Andersen, J. B. *et al.* New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Appl. Environ. Microbiol.* **64**, 2240–2246 (1998).
- Choi, S. H. & Greenberg, E. P. Genetic evidence for multimerization of LuxR, the transcriptional activator of *Vibrio fischeri* luminescence. *Mol. Mar. Biol. Biotechnol.* **6**, 408–413 (1992).
- Yates, E. A. *et al.* N-acylhomoserine lactones undergo lactonolysis in a pH-, temperature-, and acyl chain length-dependent manner during growth of *Yersinia pseudotuberculosis* and *Pseudomonas aeruginosa*. *Infect. Immun.* **70**, 5635–5646 (2002).

Supplementary Information accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank D. Karig, S. Hooshangi, S. Thiberge, M.-T. Chen and S. Subramaniam for discussions or comments on the manuscript. This material is based on work supported by the Defense Advanced Research Projects Agency (DARPA).

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to R.W. ([rweiss@princeton.edu](mailto:rweiss@princeton.edu)).

## Dynamics of *Drosophila* embryonic patterning network perturbed in space and time using microfluidics

Elena M. Lucchetta<sup>1</sup>, Ji Hwan Lee<sup>1</sup>, Lydia A. Fu<sup>1</sup>, Nipam H. Patel<sup>2</sup> & Rustem F. Ismagilov<sup>1</sup>

<sup>1</sup>Department of Chemistry, University of Chicago, Chicago, Illinois 60637, USA

<sup>2</sup>Department of Integrative Biology, Molecular and Cell Biology, and Howard Hughes Medical Institute, University of California Berkeley, Berkeley, California 94720-3140, USA

Biochemical networks are perturbed both by fluctuations in environmental conditions and genetic variation. These perturbations must be compensated for, especially when they occur during embryonic pattern formation. Complex chemical reaction networks displaying spatiotemporal dynamics have been controlled and understood by perturbing their environment in space and time<sup>1–3</sup>. Here, we apply this approach using microfluidics to investigate the robust network in *Drosophila melanogaster* that compensates for variation in the Bicoid morphogen gradient. We show that the compensation system can counteract the effects of extremely unnatural environmental conditions—a temperature step—in which the anterior and posterior halves of the embryo are developing at different temperatures and thus at different rates. Embryonic patterning was normal under this condition, suggesting that a simple reciprocal gradient system is not the mechanism of compensation. Time-specific reversals of the temperature step narrowed down the critical period for compensation to between 65 and 100 min after onset of embryonic development. The microfluidic technology used here may prove useful to future studies, as it allows spatial and temporal regulation of embryonic development.

Although rates of production and degradation of morphogens are affected by genetic and environmental variations, the mechanisms of embryo patterning have evolved to compensate for these variations. In *Drosophila melanogaster*, morphogens such as Bicoid protein act early on in the genetic hierarchy that patterns the *Drosophila* embryo along the antero-posterior axis. Altering the copy number of *bicoid* genes shifts the Bicoid expression profile and results in shifting of the expression pattern of a direct downstream target gene, *hunchback*, and subsequent expression of pair-rule and segment polarity genes<sup>4</sup>. The expression profile of the maternal Bicoid protein gradient, however, varies measurably between individual wild-type embryos, and is particularly affected by variations

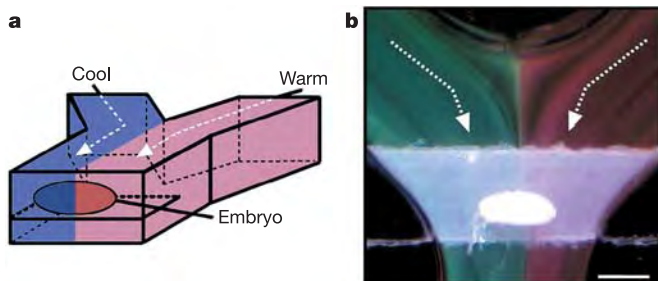
in temperature<sup>5</sup>. This naturally occurring variation does not result in altered expression of *hunchback*<sup>5</sup>, suggesting that there must be a compensatory mechanism to filter noise from the initial maternal Bicoid gradient. This compensation is important because the correct expression of gap and pair-rule genes is required to relay positional information for determining cell fate, although later compensation for shifts in patterning is also possible<sup>4</sup>. It has recently been suggested that anterior shifts in gap gene expression may be due to dynamic establishment of the Bicoid gradient in addition to regulation among the gap genes<sup>6</sup>. However, even in these analyses, the Hunchback protein boundary was stable<sup>6</sup>.

Although the exact nature of compensation is unknown, we wanted to test the limits of compensation, to determine whether compensation occurs by a simple reciprocal posterior gradient system, and to determine precisely when the compensatory mechanism functions during early development. We developed a microfluidic apparatus to precisely control the temperature of different parts of the embryo both spatially and temporally, thus differentially controlling the rate of development in the anterior and posterior halves of the embryo. In the natural world, a developing *Drosophila* embryo is unlikely to experience conditions that maintain the two halves of the embryo at different temperatures. Thermal diffusion on the scale of a *Drosophila* embryo (500  $\mu\text{m}$ ) is rapid and would equalize the temperatures in the embryo's environment and the two halves of the embryo within seconds.

We used microfluidic laminar flow<sup>7–10</sup> to create temperature differences by flowing two converging aqueous streams around an embryo, each at a controlled temperature, to provide rapid supply and removal of heat. The apparatus consists of two asymmetric moulds with alignment posts and holes<sup>11</sup>; moulds were made in polydimethylsiloxane (PDMS)<sup>12</sup>, which has low thermal conductivity. The microfluidic device can be assembled in one minute around a live embryo, with the result that the embryo is suspended in the cross-section of the channel (Fig. 1a) (see Supplementary Information for details on microfabrication, assembly of the microfluidic device and characterization of the flow).

We first demonstrated that development of the embryo was not adversely affected by flow at uniform temperature. Shear rate at the embryo was  $\sim 700 \text{ s}^{-1}$ , which was within biological range<sup>13</sup>, and dissolved oxygen in the flow was within the range of normoxia<sup>14</sup>. Embryos kept under flow in control experiments, where both streams were 20 °C or both streams were 27 °C, developed normally (see Supplementary Information).

A temperature step (*T*-step) was created by separately heating one laminar stream of flow and cooling the other. The difference in temperature was visualized using thermochromic liquid crystals<sup>15</sup> (Fig. 1b). The response time<sup>15</sup> of thermochromic liquid crystals is

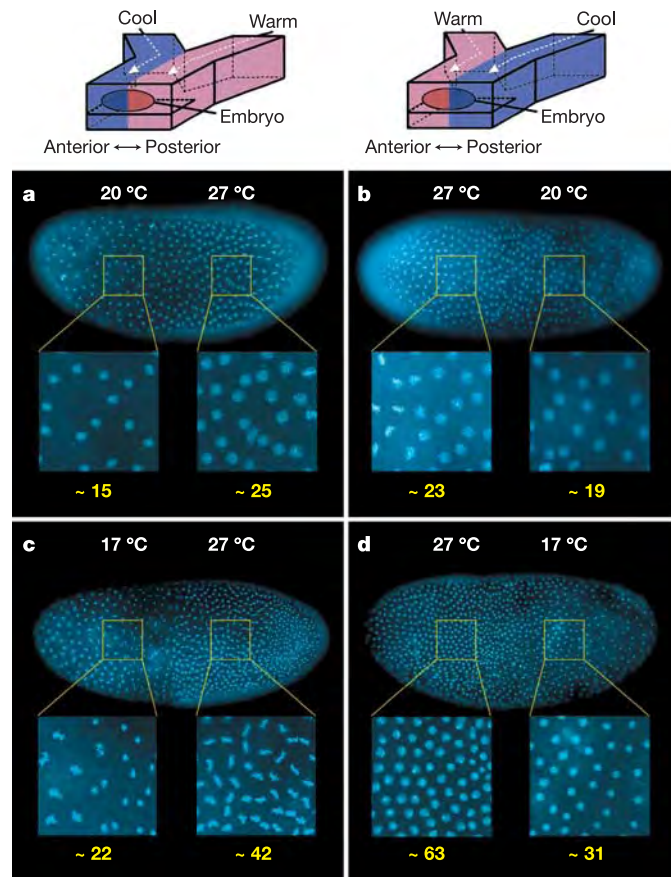


**Figure 1** Experimental set up. **a**, A schematic drawing of a PDMS microfluidic device with a *D. melanogaster* embryo developing in a temperature-step (*T*-step). **b**, A microphotograph illustrating the *T*-step around the embryo, visualized using a suspension of thermochromic liquid crystals (Image Therm Engineering) flowing at a flow rate of  $50 \text{ mm s}^{-1}$ . The temperature of the green stream is 21 °C and the temperature of the red stream is 24 °C. Scale bar, 400  $\mu\text{m}$ .

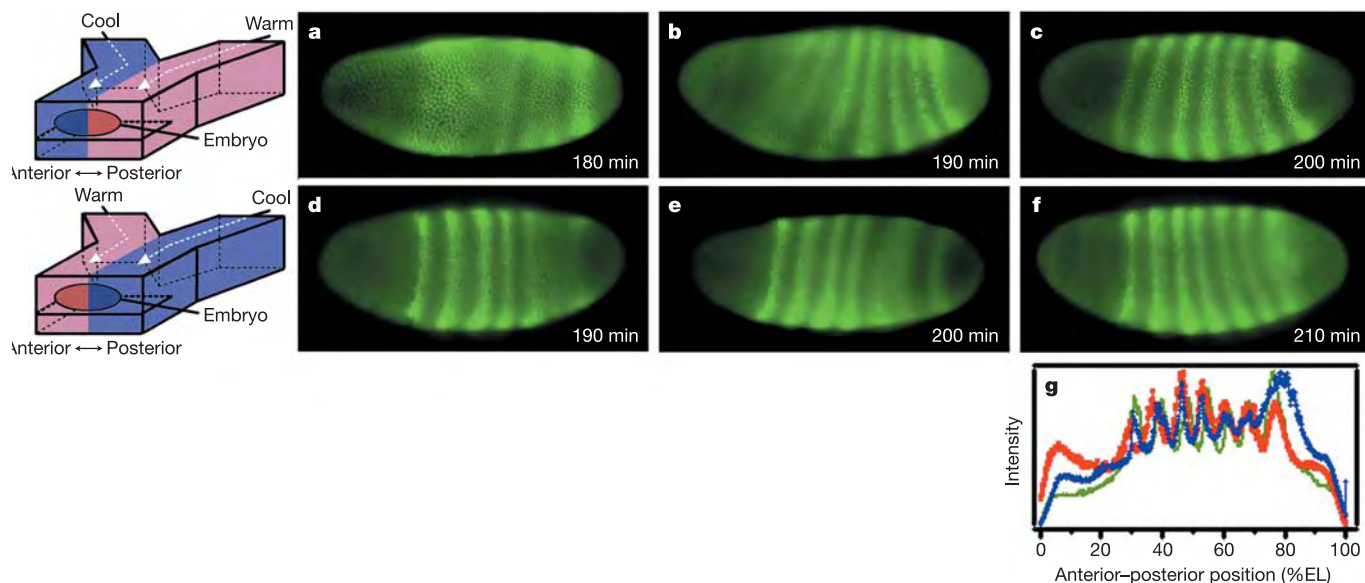
$\sim 10 \text{ ms}$ , and thus the colour at a given point in the channel is representative of the temperature 500  $\mu\text{m}$  upstream of that point. Across the embryo, thermal diffusion between the two streams was calculated to be  $\sim 50 \mu\text{m}$ , which corresponds to  $\sim 10\%$  of the egg length (EL). We calculated that the timescale of heat transfer was  $\sim 10 \mu\text{s}$  through the  $\sim 1.5\text{-}\mu\text{m}$  thick eggshell, much shorter than the timescale of heat transfer of  $\sim 1 \text{ s}$  across the  $\sim 500\text{-}\mu\text{m}$  long embryo (see Supplementary Information).

Differences in the density of nuclei in the two halves of the embryo showed experimentally that development was affected by the *T*-step. In normal embryonic development, 13 cycles of nuclear division are followed by cellularization of the embryo during the fourteenth cycle<sup>16</sup>. Although there are transient waves of mitosis, the cycles occur nearly simultaneously across the embryo and the density of nuclei stays relatively uniform.

Embryos developing in a *T*-step had a higher density of nuclei in the warmer half, confirming that the warmer half was developing more rapidly (Fig. 2). Embryos showed a greater difference when exposed to a larger *T*-step for a longer time (17 °C/27 °C for 150 min; Fig. 2c, d) than when exposed to a smaller *T*-step for a shorter time (20 °C/27 °C for 140 min; Fig. 2a, b). Embryos exposed to a *T*-step of 17 °C/27 °C showed a difference of two cell cycles (a fourfold difference in nuclear density) between the anterior and posterior halves, with the cool half closest to cycle 11 and the warm



**Figure 2** The rate of development in each half of the embryo exposed to a *T*-step is affected by temperature. Each half of the embryo is in a different cell cycle, as demonstrated by difference in nuclear density. Number of nuclei in enlarged areas shown underneath in yellow numbering. **a**, **b**, Embryos exposed to a *T*-step of 20 °C/27 °C for 140 min. **a**, Anterior half 20 °C, posterior half 27 °C. **b**, Anterior half 27 °C, posterior half 20 °C. **c**, **d**, Embryos exposed to a *T*-step of 17 °C/27 °C for 150 min. **c**, Anterior half 17 °C, posterior half 27 °C. **d**, Anterior half 27 °C, posterior half 17 °C. In all images, higher nuclear density was observed in the warmer half of the embryo.



**Figure 3** Even-skipped expression in embryos exposed to a *T*-step of 20°C/27°C. **a–c**, Embryos with a cool anterior half and warm posterior half. **d–f**, Embryos with a warm anterior half and cool posterior half. Even-skipped stripes were consistently expressed in the warm half first (**a** and **b**, **d** and **e**), but resolved in the correct positions (**c**, **f**).

**g**, Intensity profile of Even-skipped expression in embryos exposed to the *T*-step in panels **c** (red) and **f** (blue), compared to an average intensity profile (green) of Even-skipped expression in four control embryos that developed at room temperature.

half closest to cycle 13 in control embryos. Therefore, the difference in nuclear density was more than a mitotic wave. Embryos exposed to a *T*-step of 20°C/27°C showed a difference of one cell cycle (a twofold difference in nuclear density) between the anterior and posterior halves, with the cool half and warm half closest to cycle 11 and 12 in controls, respectively (see Supplementary Information). Notably, the boundary in nuclear density was relatively sharp, not only confirming a tight thermal boundary, but demonstrating that cell cycle regulation could occur fairly independently between neighbouring nuclei in the syncytial blastoderm. Wasp embryos set in an agar block over heating elements have previously been exposed to temperature gradients<sup>17</sup>. Notably, these experiments resulted in normal patterning late in development, but as no molecular markers were assayed, it was unclear at what point compensation occurred. Furthermore, the morphogen gradients involved in patterning in wasps are unknown.

Our initial presumption was that *Drosophila* embryos developing in the unusual environment established within our microfluidic apparatus would have obvious defects in patterning. It has been shown that the shape of the Bicoid gradient, which is determined by a combination of production, degradation and diffusion of Bicoid, is strongly affected by temperature<sup>5</sup>. In our experiments, these processes will occur at different rates in the two halves of embryos. To our surprise, embryos allowed to develop in a temperature step (either 17°C/27°C or 20°C/27°C) for 150 min and then left at room temperature developed into normal larvae with the correct number and pattern of segments (data not shown). The normal appearance of larvae suggests that the embryo compensates for the unnatural temperature environment.

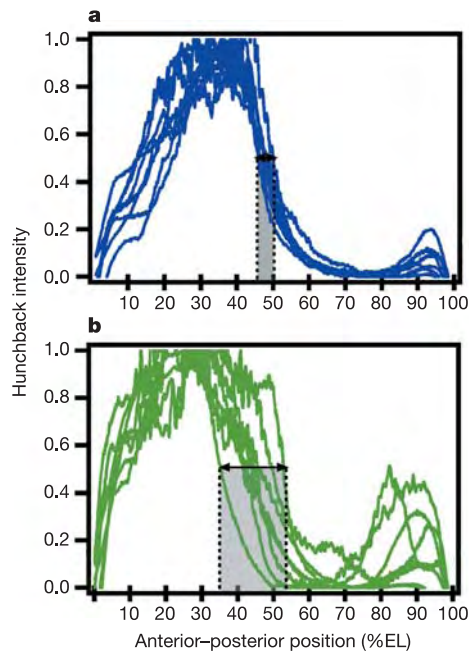
We used the microfluidic system to test whether patterning of Even-skipped<sup>18</sup> remained robust in embryos exposed to the *T*-step. Normally, Even-skipped resolves into seven stripes in a specific order. First, stripes 1 and 2 resolve, and then stripe 7 resolves. Stripes 3, 4, 5, and 6 resolve next, more or less simultaneously<sup>18</sup>. Once resolved (by late cycle 14), the stripes are positioned at ~31, 40, 48, 55, 62, 69, and 78% EL<sup>19</sup>. The pattern of Even-skipped expression was monitored at the *T*-step of 20°C/27°C (Fig. 3), which was chosen so that we could observe the entire embryo within cycle 14. A temperature difference greater than this causes gastrulation and

germband extension to begin in the warmer half before Even-skipped stripes in the cooler half resolve (see Supplementary Information), making measurement of stripe positions inaccurate, but illustrating that the two halves still develop at different rates. Embryos exposed to the *T*-step (20°C/27°C) clearly formed Even-skipped stripes out of order. When the anterior half of the embryo was warmer, Even-skipped stripes 1, 2 and 3 consistently resolved first (Fig. 3d, e). When the posterior half of the embryo was warmer, Even-skipped stripes 5, 6 and 7 consistently resolved first (Fig. 3a, b); this was particularly striking because in normal development stripes 5 and 6 resolve last.

Surprisingly, Even-skipped stripe positions were precise, despite the difference in developmental rate in the two halves of the embryo (Fig. 3c, f). Even-skipped stripes 1 through 7 resolved at 30, 38, 48, 54, 62, 68, and 78 ± 1% EL, respectively, for embryos exposed to a *T*-step of 20°C/27°C (with either half being warmer) (Fig. 3g). Embryos exposed to the *T*-step of 20°C/27°C also compensated for the unnatural environment at the level of Hunchback expression, which is a more direct readout of the maternal Bicoid morphogen gradient. For embryos with anterior 27°C/posterior 20°C (Fig. 4a) and embryos with anterior 20°C/posterior 27°C (data not shown), the Hunchback boundary remained spatially precise (46–51% EL), comparable to control embryos<sup>5</sup>.

We determined the critical period during which the compensation mechanism functions by setting up a temperature regimen as follows: embryos developed with anterior at 27°C/posterior at 20°C with the exception of a brief 35-min period during which the orientation of the *T*-step was transiently reversed (anterior 20°C/posterior 27°C at  $t_1$ ) before being reversed back (anterior 27°C/posterior 20°C at  $t_2$ ). Embryos exposed to a brief temperature reversal ( $t_1$  = 65 min and  $t_2$  = 100 min) showed a marked increase in variability of the Hunchback boundary (35–53% EL) with anterior bias (Fig. 4b), resembling the embryos carrying the *stau*<sup>HL</sup> or the *stau*<sup>r9</sup> allele<sup>5</sup>, which also showed variability with anterior bias. Corresponding nuclear images stained with 4,6-diamidino-2-phenylindole (DAPI) do not support the possibility that the increased variability of the Hunchback boundary position is simply due to differences in the ages of the embryos in cycle 14 (see Supplementary Information). Other earlier ( $t_1$  = 35 min,





**Figure 4** Hunchback expression in embryos exposed to a *T*-step and time-dependent *T*-step. **a**, Hunchback intensity profiles of embryos with anterior half at 27 °C and posterior half at 20 °C. Hunchback position was normal, varying over 5% EL (46–51% EL). **b**, Hunchback intensity profiles of embryos exposed to a time-dependent *T*-step with anterior half at 27 °C and posterior half at 20 °C, with the exception of a brief temperature reversal (anterior 20 °C/posterior 27 °C) between 65 and 100 min. The position of Hunchback was variable over 18% EL (35–53% EL).

$t_2 = 70$  min) or later ( $t_1 = 95$  min,  $t_2 = 130$  min) transient reversals of the *T*-step resulted in relatively low variability in the Hunchback boundary (42–49% and 43–50% EL, respectively) (see Supplementary Information). These results show that a key part of the compensatory mechanism is established between 65 and 100 min. By extension, we suggest that normally occurring variations are ‘measured’ and possibly compensated for during this stage of development.

To test the possible polarity of the correction mechanism, we inverted the temperature regimen: embryos developed with anterior at 20 °C and posterior at 27 °C, with the exception of a brief 35-min period during which the orientation of the *T*-step was transiently reversed (anterior 27 °C/posterior 20 °C at  $t_1$ ) before being reversed back (anterior 20 °C/posterior 27 °C at  $t_2$ ). Embryos exposed to later temperature reversals ( $t_1 = 65$  min,  $t_2 = 100$  min or  $t_1 = 95$  min,  $t_2 = 130$  min) showed low variability in the Hunchback boundary (43–48% or 42–49% EL, respectively). Embryos exposed to an earlier temperature reversal ( $t_1 = 35$  min,  $t_2 = 70$  min) also showed low variability in the Hunchback boundary (51–55% EL), but in this case with a slight posterior bias (see Supplementary Information).

Using microfluidics, we changed the relative rate of development of the two halves of the embryo. Interestingly, Even-skipped expression in Fig. 3d and e bears resemblance to the pattern seen in beetle embryos<sup>20</sup>, in that Even-skipped stripes do not all appear more or less simultaneously. Despite our ability to change the temporal order in which Even-skipped stripes form, the spatial precision of the pattern remained intact. Thus, the low variability of the Hunchback boundary and Even-skipped pattern in these embryos suggests that the compensatory mechanism of the embryo can allow for correct patterning even in extremely unnatural environments.

A plausible compensatory mechanism for maintaining a stable

Hunchback boundary during normal development could involve an opposing gradient system using the anterior Bicoid gradient plus an unknown posterior gradient. Previous work has ruled out many genetic candidates for the possible posterior gradient (including *nanos* and *caudal*), but found that certain *staufer* alleles affected precision, suggesting that the compensatory system might involve the localization of an uncharacterized messenger RNA<sup>5</sup>. Our results suggest that the compensation system is not simply a reciprocal posterior gradient, as such a system would have failed when the two halves of the embryo developed at two different temperatures. Temperature reversal experiments lead us to suggest that the compensation mechanism has a critical establishment phase corresponding to 65–100 min of development.

Understanding the dynamics of a complex system by perturbing its environment in space and time does not require *a priori* knowledge of the system’s components. This approach has been used to understand and control biological systems (such as appearance of fibrillations in a mammalian heart<sup>21</sup>) and chemical reaction networks (such as formation of patterns in the Belousov-Zhabotinsky reaction<sup>3</sup> and in surface-catalysed CO oxidations<sup>1</sup>). We believe that the microfluidic methods we have established here will prove to be an experimentally powerful approach that will allow the environment of the embryos to be spatiotemporally controlled. Perturbing the environment is a complementary approach to perturbing the molecular components of the network, as it might provide information on where and when events occur, rather than which molecules are involved. A combination of these two approaches might prove especially useful for identifying correction mechanisms responsible for the robustness of developmental and other biochemical networks. □

## Methods

### Fly stocks, fabrication of the microfluidic device, and experimental set up

All embryos were collected using wild-type Oregon R. *Drosophila*. Multilevel PDMS moulds were made using rapid prototyping. Embryos were collected at 23 °C for 2–5 min, placed on double-sided tape and assembled in the microfluidic device. Embryos were exposed to flow within 10 min of the start of collection. Syringe pumps (KdScientific) were used to control volumetric flow rate (2 ml min<sup>−1</sup> total). Temperature of each laminar stream was maintained using a double chilling/heating plate (Echo therm, Torrey Pines Scientific).

### Immunostaining

After removal from flow, embryos were dechorionated, fixed in 3% formaldehyde in PEM buffer and immunostained using standard methods with anti-Even-skipped (mouse monoclonal 2B8) and anti-Hunchback (mouse monoclonal 1G10) antibodies, and goat anti-mouse IgG (H + L) AlexaFluor 488 conjugated secondary antibody (Molecular Probes).

### Image acquisition and analysis

Images were acquired using a Leica DM IRE2 inverted microscope with a  $\times 20$  0.7 NA objective and cooled CCD camera ORCA ERG 1394 (12-bit, 1344  $\times$  1024 resolution, Hamamatsu Photonics). Even-skipped and Hunchback expression intensity profiles were analysed using MetaMorph Imaging System (Universal Imaging Corp).

See Supplementary Information for details regarding Methods.

Received 11 November 2004; accepted 23 February 2005; doi:10.1038/nature03509.

- Wolff, J. *et al.* Spatiotemporal addressing of surface activity. *Science* **294**, 134–137 (2001).
- Sakurai, T., Mihailuk, E., Chirila, F. & Showalter, K. Design and control of wave propagation patterns in excitable media. *Science* **296**, 2009–2012 (2002).
- Vanag, V. K. *et al.* Oscillatory cluster patterns in a homogeneous chemical system with global feedback. *Nature* **406**, 389–391 (2000).
- Driever, W. & Nusslein-Volhard, C. The Bicoid protein determines position in the *Drosophila* embryo in a concentration-dependent manner. *Cell* **54**, 95–104 (1988).
- Houchmandzadeh, B., Wieschaus, E. & Leibler, S. Establishment of developmental precision and proportions in the early *Drosophila* embryo. *Nature* **415**, 798–802 (2002).
- Jaeger, J. *et al.* Dynamic control of positional information in the early *Drosophila* embryo. *Nature* **430**, 368–371 (2004).
- Kenis, P. J. A., Ismagilov, R. F. & Whitesides, G. M. Microfabrication inside capillaries using multiphase laminar flow patterning. *Science* **285**, 83–85 (1999).
- Takayama, S. *et al.* Subcellular positioning of small molecules. *Nature* **411**, 1016 (2001).
- Hatch, A. *et al.* A rapid diffusion immunoassay in a T-sensor. *Nature Biotechnol.* **19**, 461–465 (2001).
- Sia, S. K. & Whitesides, G. M. Microfluidic devices fabricated in poly(dimethylsiloxane) for biological studies. *Electrophoresis* **24**, 3563–3576 (2003).
- Anderson, J. R. *et al.* Fabrication of topologically complex three-dimensional microfluidic systems in PDMS by rapid prototyping. *Anal. Chem.* **72**, 3158–3164 (2000).

12. McDonald, J. C. & Whitesides, G. M. Poly(dimethylsiloxane) as a material for fabricating microfluidic devices. *Accounts Chem. Res.* **35**, 491–499 (2002).
13. Furlong, E. E. M., Proffitt, D. & Scott, M. P. Automated sorting of live transgenic embryos. *Nature Biotechnol.* **19**, 153–156 (2001).
14. DiGregorio, P. J., Ubersax, J. A. & O'Farrell, P. H. Hypoxia and nitric oxide induce a rapid, reversible cell cycle arrest of the *Drosophila* syncytial divisions. *J. Biol. Chem.* **276**, 1930–1937 (2001).
15. Stasiek, J. A. & Kowalewski, T. A. Thermochromic liquid crystals applied for heat transfer research. *Opto-Electron. Rev.* **10**, 1–10 (2002).
16. Foe, V. E., Odell, G. M. & Edgar, B. A. in *The Development of Drosophila melanogaster* (eds Bate, M. & Martinez Arias, A.) (Cold Spring Harbor Press, New York, 1993).
17. Niemuth, J. & Wolf, R. Developmental asynchrony caused by steep temperature-gradients does not impair pattern-formation in the wasp, *Pimpla turionellae* L. *Roux's Arch. Dev. Biol.* **204**, 444–452 (1995).
18. Frasch, M. et al. Characterization and localization of the Even-skipped protein of *Drosophila*. *EMBO J.* **6**, 749–759 (1987).
19. Myasnikova, E. et al. Registration of the expression patterns of *Drosophila* segmentation genes by two independent methods. *Bioinformatics* **17**, 3–12 (2001).
20. Patel, N. H., Condrón, B. G. & Zinn, K. Pair-rule expression patterns of Even-skipped are found in both short-germ and long-germ beetles. *Nature* **367**, 429–434 (1994).
21. Witkowski, F. X. et al. Spatiotemporal evolution of ventricular fibrillation. *Nature* **392**, 78–82 (1998).

**Supplementary Information** accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** This work was supported by the Searle Scholars Program and was performed at the Chicago MRSEC microfluidic facility funded by the NSF. N.H.P. is an Investigator of the Howard Hughes Medical Institute. We thank J. B. Brokaw, C. A. Macrandner and M. Giorgianni for preliminary experiments. We thank D. Bilder and I. Hariharan for discussions and comments on the manuscript.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to R.F.I. (r-ismagilov@uchicago.edu) and N.H.P. (nipam@uclink.berkeley.edu).

## IKK $\alpha$ limits macrophage NF- $\kappa$ B activation and contributes to the resolution of inflammation

Toby Lawrence<sup>1\*†</sup>, Magali Bebieñ<sup>1\*</sup>, George Y. Liu<sup>2</sup>, Victor Nizet<sup>2</sup> & Michael Karin<sup>1</sup>

<sup>1</sup>Laboratory of Gene Regulation and Signal Transduction, Department of Pharmacology, and <sup>2</sup>Division of Paediatric Infectious Diseases, School of Medicine, University of California San Diego, 9500 Gilman Drive, California 92093, USA

\* These authors contributed equally to this work

† Present address: Kennedy Institute of Rheumatology Division, Faculty of Medicine, Imperial College, 1 Aspenlea Road, London W6 8LH, UK

Inflammation and innate immunity involve signalling pathways leading to the production of inflammatory mediators. Usually such responses are self-limiting, but aberrant resolution of inflammation results in chronic diseases<sup>1</sup>. Much attention has focused on pro-inflammatory signalling but little is known about the mechanisms that resolve inflammation. The I $\kappa$ B kinase (IKK) complex contains two catalytic subunits, IKK $\alpha$  and IKK $\beta$ , and controls the activation of NF- $\kappa$ B transcription factors, which play a pivotal role in inflammation<sup>2</sup>. Ample evidence indicates that IKK $\beta$  mediates NF- $\kappa$ B activation in response to pro-inflammatory cytokines and microbial products. IKK $\alpha$  regulates an alternative pathway important for lymphoid organogenesis<sup>3</sup>, but the role of IKK $\alpha$  in inflammation is unknown. Here we describe a new role for IKK $\alpha$  in the negative regulation of macrophage activation and inflammation. IKK $\alpha$  contributes to suppression of NF- $\kappa$ B activity by accelerating both the turnover of the NF- $\kappa$ B subunits RelA and c-Rel, and their removal from

pro-inflammatory gene promoters. Inactivation of IKK $\alpha$  in mice enhances inflammation and bacterial clearance. Hence, the two IKK catalytic subunits have evolved opposing but complimentary roles needed for the intricate control of inflammation and innate immunity.

NF- $\kappa$ B transcription factors are pivotal regulators of inflammation and immunity that control expression of important immunoregulatory genes<sup>2,3</sup>. NF- $\kappa$ B activation and activity are tightly controlled by a number of endogenous mechanisms that limit the excessive and prolonged production of pro-inflammatory mediators, which can cause tissue damage during the inflammatory response<sup>3,4</sup>. With the exception of autoregulated I $\kappa$ B $\alpha$  (inhibitor of NF- $\kappa$ B alpha) expression<sup>4</sup> and induction of the de-ubiquitinating enzyme A20 (ref. 5), the mechanisms that limit the duration and magnitude of NF- $\kappa$ B signalling are poorly understood. It is likely that I $\kappa$ B $\alpha$  and A20 are not the only physiologically relevant negative regulators of this central signalling module. The activation of NF- $\kappa$ B by pro-inflammatory stimuli depends on the classical IKK complex, composed of two catalytic subunits (IKK $\alpha$  and IKK $\beta$ ), together with a regulatory subunit IKK $\gamma$ /NEMO<sup>3,6</sup>. IKK activation is triggered by engagement of cytokine receptors as well as pattern recognition receptors. Gene disruption studies revealed that in addition to IKK $\gamma$ /NEMO, which is necessary for activation of the classical IKK complex<sup>7</sup>, it is IKK $\beta$  rather than IKK $\alpha$  that plays a more critical role in activating inflammation<sup>6</sup>. IKK $\alpha$  forms an alternative complex (without IKK $\beta$  and IKK $\gamma$ )<sup>8</sup>, the function of which is required for lymphoid organ development and B cell maturation<sup>9</sup>. This alternative signalling pathway is activated by certain members of the tumour necrosis factor (TNF) family, but not by pattern recognition receptors such as Toll-like receptor 4 (TLR4) (ref. 2). The function of IKK $\alpha$  within the classical IKK complex, however, is not entirely clear. Although a chromatin-modifying function for IKK $\alpha$  required for TNF $\alpha$ -mediated gene induction has been suggested<sup>10,11</sup>, targeting of the *Ikk $\alpha$*  (also called *Chuk*) gene in mice does not support this proposal<sup>2</sup>.

We investigated the role of IKK $\alpha$  in inflammation and innate immunity *in vivo*, using mice that express the inactivatable variant IKK $\alpha$ (AA) (ref. 9). *Ikk $\alpha$* <sup>AA/AA</sup> mice (which are homozygous for the mutant allele) and littermate controls were challenged systemically with the Gram-positive human pathogen group B *Streptococcus* (GBS)<sup>12</sup>, and monitored for bacterial clearance and survival. Although *Ikk $\alpha$* <sup>AA/AA</sup> mice showed significantly decreased blood bacterial counts at 24 h (Fig. 1a), mortality was notably accelerated relative to wild-type animals (Fig. 1b). This paradoxical result suggested that *Ikk $\alpha$* <sup>AA/AA</sup> mice have an exacerbated inflammatory response to infection that enhances bacterial clearance but provokes septic shock. To pursue this hypothesis further, we assessed the local inflammatory response to bacterial infection in a non-lethal model. When *Ikk $\alpha$* <sup>AA/AA</sup> mice and littermate controls were inoculated intranasally with GBS, the mutants showed increased bacterial clearance, associated with increased neutrophil recruitment and local inflammation (Fig. 1c–e). Together, these studies indicate that IKK $\alpha$  is somehow involved in limiting the inflammatory response to Gram-positive infection.

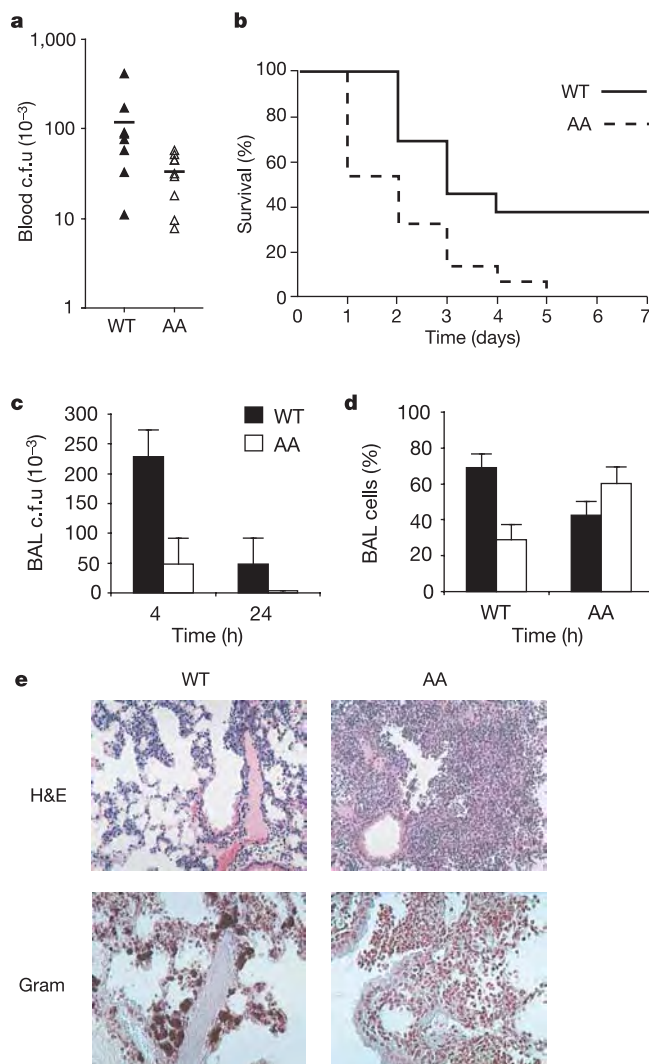
We extended these studies to a model of Gram-negative septic shock by using the TLR4 agonist lipopolysaccharide (LPS) from *Escherichia coli*<sup>13</sup>. *Ikk $\alpha$* <sup>AA/AA</sup> mice showed increased susceptibility to LPS-induced septic shock (Fig. 2a). Real-time quantitative polymerase chain reaction (PCR) analysis of liver and lung RNA showed elevated expression of pro-inflammatory and anti-apoptotic NF- $\kappa$ B target genes, including macrophage inflammatory protein (MIP)-2, MIP-1 $\alpha$ , interleukin (IL)-12p40, inhibitor of apoptosis protein (IAP)-2 and inducible nitric oxide synthase (iNOS) (Fig. 2b). Local LPS administration to the lung also resulted in elevated pulmonary leukocyte recruitment and pro-inflammatory cytokine production in *Ikk $\alpha$* <sup>AA/AA</sup> mice relative to wild-type littermates (see Supplementary Fig. 1). Neutrophil

recruitment in zymosan A-induced peritonitis (a TLR2-dependent stimulus<sup>13,14</sup>) was similarly exacerbated in *Ikkα<sup>AA/AA</sup>* mice, in association with increased chemokine and IL-12 release (see Supplementary Fig. 2). These findings raised the possibility that IKKα might serve to suppress or terminate activation of the classical NF-κB pathway during TLR signalling.

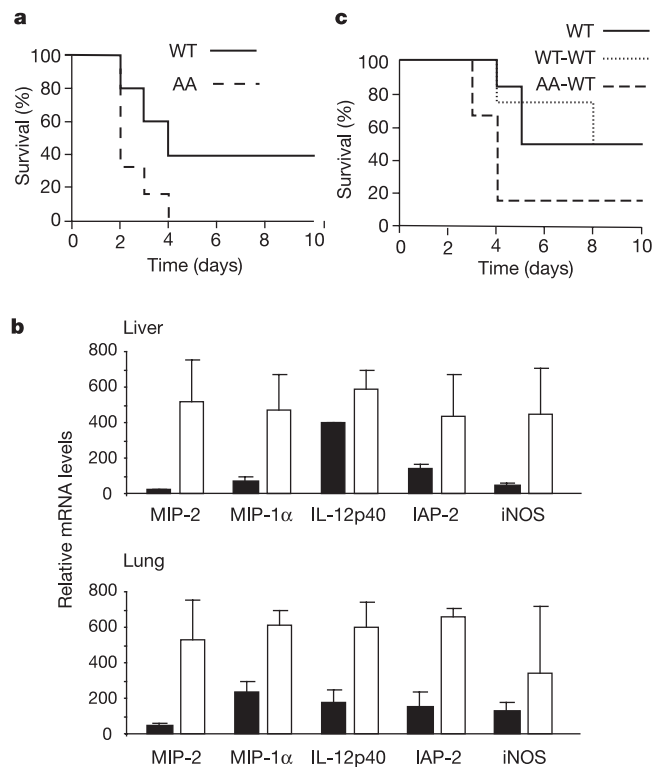
Further experiments demonstrated that the exacerbated inflammatory phenotype was intrinsic to the haematopoietic system. Lethally irradiated wild-type mice were reconstituted with bone marrow from *Ikkα<sup>AA/AA</sup>* mice or littermate controls. Mice receiving *Ikkα<sup>AA/AA</sup>* bone marrow were more susceptible to LPS-induced septic shock (Fig. 2c). Radiation chimaeras were also infected with GBS intranasally. Mice receiving *Ikkα<sup>AA/AA</sup>* bone marrow showed enhanced bacterial clearance (T. Lawrence, unpublished observations). These results are in contrast with the role of IKKα in

lymphoid organogenesis, which is due to an intrinsic defect in lymphotoxin-β signalling in radiation-resistant stromal cells<sup>15</sup>, and suggest a distinct function for IKKα in the resolution of inflammation in macrophages or other haematopoietic cells expressing TLRs.

Macrophages have a key role in innate immunity and inflammation<sup>16</sup>. Considering that enhanced inflammation in *Ikkα<sup>AA/AA</sup>* mice was intrinsic to the haematopoietic compartment, we examined the response of primary macrophages to bacteria and TLR agonists. Alveolar and peritoneal macrophages from *Ikkα<sup>AA/AA</sup>* mice show dramatically increased bactericidal activity against GBS *in vitro* (Fig. 3a). An important aspect of GBS evasion of the immune system is the induction of macrophage apoptosis<sup>12</sup>, a response that is suppressed in similar models by IKKβ-driven NF-κB activation<sup>17</sup>. We found that *Ikkα<sup>AA/AA</sup>* macrophages were more resistant to GBS-induced apoptosis *in vitro* and *in vivo* compared with their wild-type counterparts (Fig. 3b). Real-time PCR analysis revealed increased induction of messenger RNAs for NF-κB-dependent anti-apoptotic and pro-inflammatory genes in GBS-infected *Ikkα<sup>AA/AA</sup>* macrophages (Fig. 3c). *In vitro*, *Ikkα<sup>AA/AA</sup>* macrophage resistance to GBS-induced cell death was associated with sustained iNOS expression and nitric oxide production (Supplementary Fig. 3), which is an important antimicrobial mechanism against GBS infection<sup>18</sup>. These effects of IKKα inactivation were mimicked by the addition of exogenous granulocyte-macrophage colony-stimulating factor (GM-CSF), a macrophage survival factor, to wild-type cells (Supplementary Fig. 3). We suggest that aberrant regulation of apoptotic pathways, leading to protracted survival of



**Figure 1** IKKα limits the inflammatory response to Gram-positive infection. **a**, GBS titres in blood were measured after intravenous inoculation with  $5 \times 10^6$  c.f.u. bacteria. Data are represented as mean c.f.u. from individual wild-type (WT, closed triangles) and *Ikkα<sup>AA/AA</sup>* (AA, open triangles) mice, median is indicated by a solid bar. **b**, Kaplan–Meier survival plot of AA and WT littermates ( $n = 11$ ). **c**, GBS-induced pneumonia in WT and AA mice, represented as bacterial titres in bronchial-alveolar lavage (BAL) ( $n = 6–8$ ). **d**, Differential cell counts of bronchial-alveolar lavage, with data showing percentages of neutrophils (open bars) and macrophages (filled bars). **e**, Histopathology of GBS-induced pneumonia in WT and AA mice. Gram and haematoxylin-eosin staining (H&E) of lung tissue (original magnification  $\times 400$ ). Data in **c** and **d** show mean  $\pm$  s.d. ( $n = 6–8$ ).



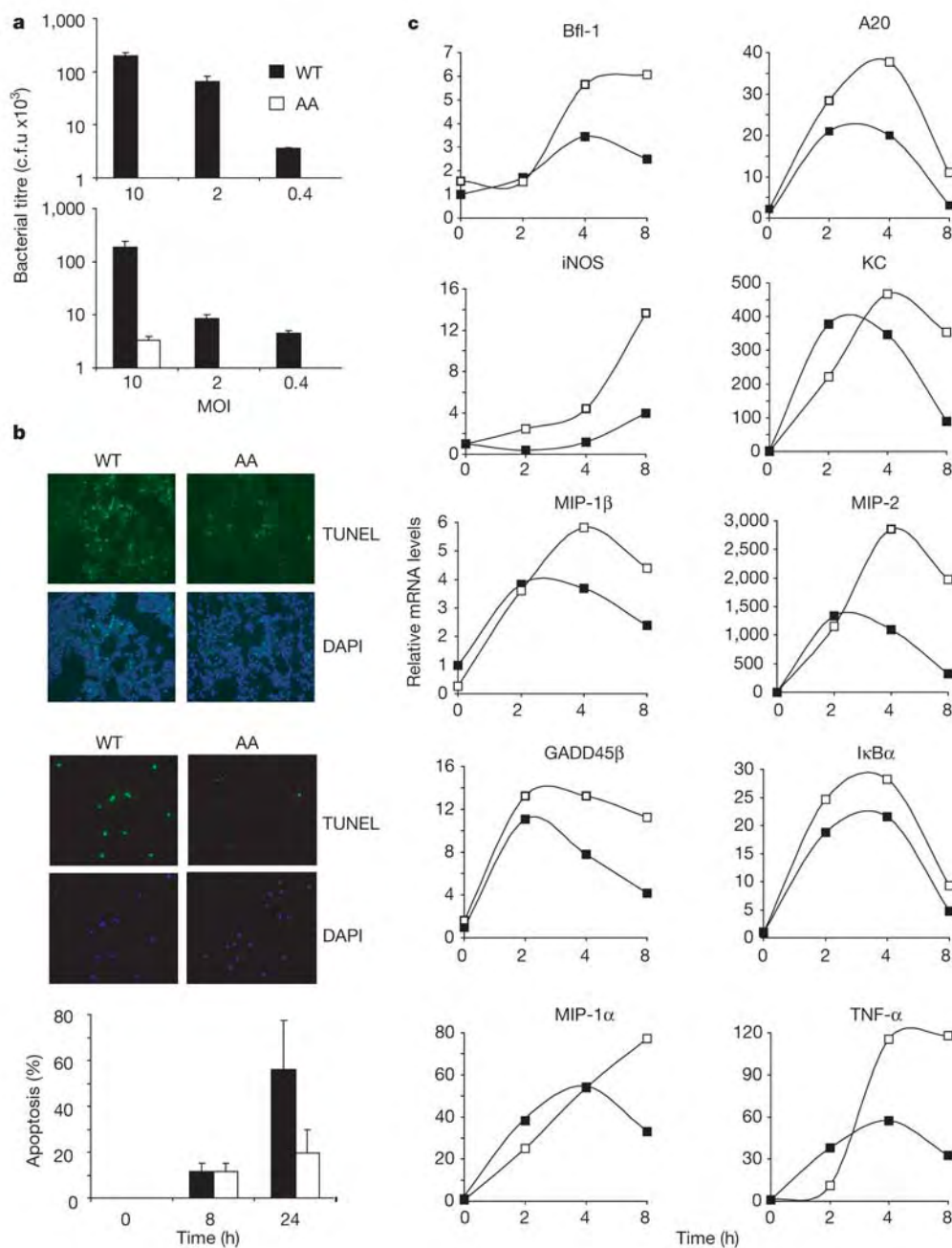
**Figure 2** IKKα deficiency in haematopoietic cells increases the systemic inflammatory response to LPS. **a**, Kaplan–Meier survival plot for WT and AA mice ( $n = 6–8$ ) after intraperitoneal injection of *E. coli* LPS. **b**, Lung and liver were harvested 24 h after LPS injection and RNA was isolated for real-time PCR analysis. Data are mean relative mRNA levels in WT (filled bars) and AA (open bars) mice, normalized to cyclophilin mRNA expression and shown as mean  $\pm$  s.d. ( $n = 3–4$ ). **c**, Kaplan–Meier survival plots for radiation chimaera mice. Wild-type mice were lethally irradiated and reconstituted with WT (WT-WT) or AA (WT-AA) bone marrow. After 8 weeks, mice were challenged with LPS.



activated macrophages, may in part explain the phenotype of enhanced inflammation and bactericidal function in *Ikk $\alpha$ <sup>AA/AA</sup>* mice<sup>12</sup>.

Further studies using real-time PCR showed increased pro-inflammatory and anti-apoptotic gene expression in *Ikk $\alpha$ <sup>AA/AA</sup>* macrophages challenged with LPS *in vitro* (Fig. 4a). Elevated expression of such genes was confirmed by ribonuclease (RNase) protection assays (Fig. 4b), indicating that the major effect of the mutation was increased duration of NF- $\kappa$ B-dependent gene

induction. Accordingly, the initial induction of NF- $\kappa$ B DNA-binding activity was not significantly elevated in LPS-stimulated *Ikk $\alpha$ <sup>AA/AA</sup>* macrophages, but at later time points NF- $\kappa$ B DNA-binding activity was increased in the mutant cells relative to wild-type controls (Fig. 4c). Notably, increased nuclear levels of the NF- $\kappa$ B subunits RelA and c-Rel were detected at later time points in *Ikk $\alpha$ <sup>AA/AA</sup>* relative to wild-type macrophages (Fig. 4c). Chromatin immunoprecipitation (ChIP) assays showed that the duration of RelA and c-Rel residence at the Bcl-2 family member A1 (Bfl-1),



**Figure 3** *IKK $\alpha$*  suppresses macrophage activation in response to GBS infection *in vitro*. **a**, Alveolar macrophages (top panel) and peritoneal macrophages (bottom panel) from WT and AA mice were infected with GBS, and intracellular titres of bacteria were determined 4 h later. Very low bacterial titres were recovered from AA macrophages. **b**, Top panel, lung tissue sections from WT and AA mice were examined by TUNEL assay after intranasal infection with GBS. Middle panel, TUNEL assay on peritoneal macrophages after infection with GBS at an MOI = 10:1. Bottom panel, quantification of

apoptosis in GBS-infected macrophages (WT, filled bars; AA, open bars). Data in **a** and **b** show mean  $\pm$  s.d. ( $n = 3$ ). **c**, RNA was isolated from GBS-infected peritoneal macrophages, and gene expression was quantified by real-time PCR and normalized to the level of cyclophilin mRNA. WT, filled squares; AA, open squares. KC, mouse homologue of human melanoma growth stimulatory activity (MGSA); GADD45 $\beta$ , growth arrest DNA damage-inducible gene 45 $\beta$ .

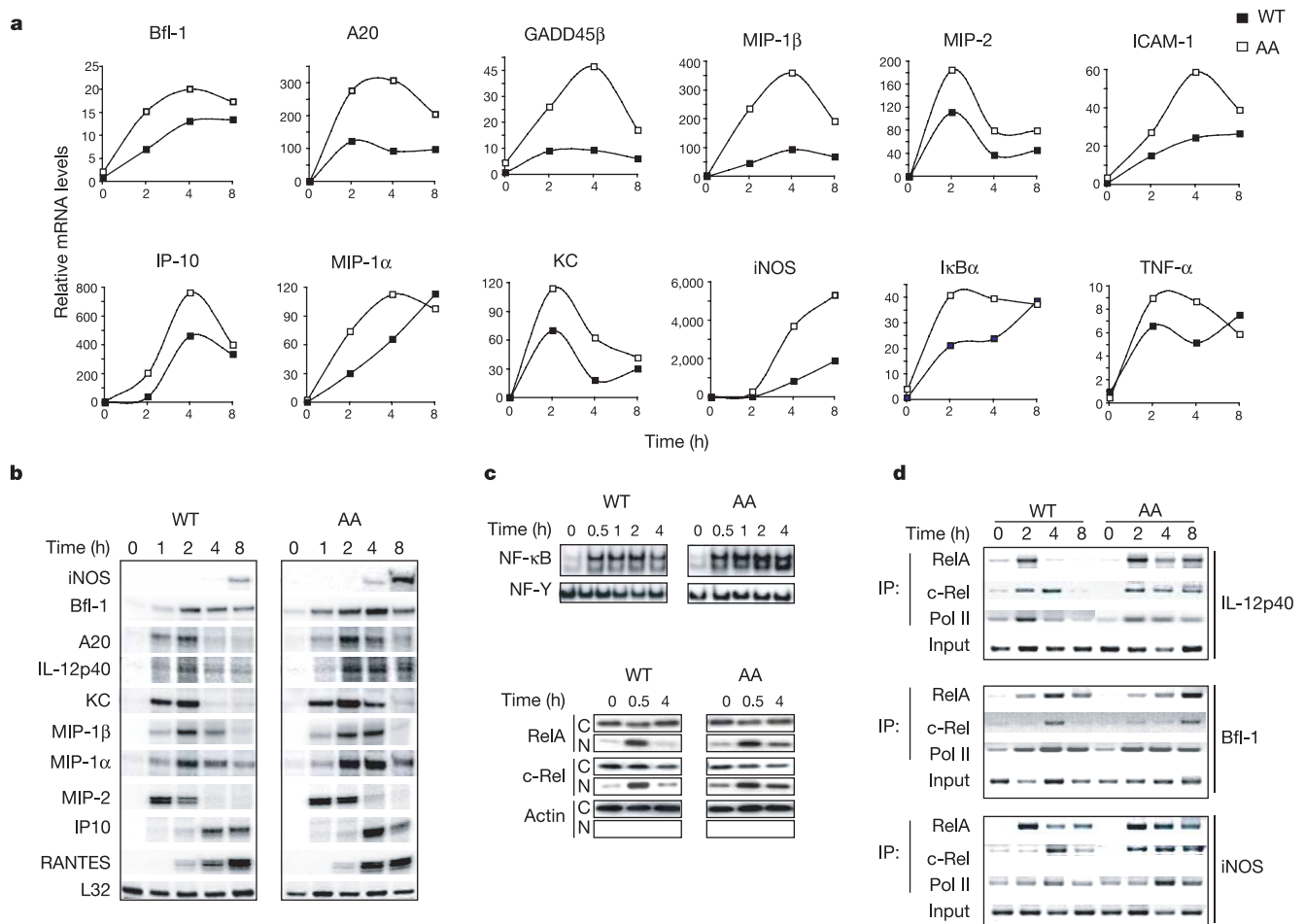
IL-12p40 and iNOS gene promoters was extended in LPS-stimulated *Ikk $\alpha$ <sup>AA/AA</sup>* macrophages relative to wild-type cells (Fig. 4d). Collectively, these results suggest that IKK $\alpha$  activation might be required for accelerated promoter clearance of RelA- or c-Rel-containing NF- $\kappa$ B complexes during resolution of the inflammatory response.

Despite the changes described above, the kinetics of IKK $\beta$ -dependent I $\kappa$ B $\alpha$  phosphorylation and degradation in LPS-stimulated macrophages were not affected by the *Ikk $\alpha$ <sup>AA</sup>* mutation (Fig. 5a). The IKK complex is suggested to phosphorylate the carboxy-terminal activation domain of RelA<sup>19</sup>, but the physiological function of this phosphorylation and the IKK subunit that mediates it are unknown. In contrast to the normal levels of I $\kappa$ B $\alpha$  kinase activity, *Ikk $\alpha$ <sup>AA/AA</sup>* macrophages showed reduced activation of an IKK $\gamma$ -associated RelA kinase (Fig. 5b). The amino acid target for this kinase activity is likely to be serine 536, because a variant of RelA in which this serine was replaced with alanine was phosphorylated equally well by IKK complexes from wild-type or *Ikk $\alpha$ <sup>AA/AA</sup>* macrophages (see Supplementary Fig. 4a). In similar experiments, IKK complexes from LPS-stimulated macrophages did not phosphorylate an amino-terminal fragment of RelA (amino acids 1–305) (data not shown). Immunoblot analysis with phospho-specific antibodies that recognize RelA phosphorylated at either S536 or S276 showed diminished S536 phosphorylation in LPS-stimulated

*Ikk $\alpha$ <sup>AA/AA</sup>* macrophages, but S276 phosphorylation remained intact (Fig. 5c). c-Rel is also subjected to C-terminal phosphorylation<sup>20,21</sup>, and the C-terminal kinase activity of c-Rel was also decreased in IKK complexes from *Ikk $\alpha$ <sup>AA/AA</sup>* macrophages, provided that the c-Rel fragment used in these assays was restricted to residues 422–540; phosphorylation of a longer c-Rel substrate encompassing residues 422–588 was not as extensively reduced (see Supplementary Fig. 4b).

Promoter clearance of sequence-specific transcription factors is suggested to be mediated by ubiquitin-dependent proteolysis<sup>22–24</sup>, which is closely linked to inducible phosphorylation and ubiquitination of transcription factor trans-activation domains<sup>22</sup>. We therefore examined whether IKK $\alpha$  was involved in Rel protein turnover during the course of macrophage activation. Previous studies have demonstrated ubiquitination and proteasomal degradation of both RelA and c-Rel<sup>21,24–26</sup>, and it has been suggested that RelA ubiquitination leads to recruitment of proteasome components to target gene promoters and to RelA proteolysis<sup>24</sup>. This mechanism was proposed to be involved in the termination of NF- $\kappa$ B activation.

We stimulated primary macrophages with LPS and followed the kinetics of NF- $\kappa$ B activation in the absence or presence of a proteasome inhibitor. Proteasome inhibition prolonged the LPS-induced DNA-binding activity of NF- $\kappa$ B and increased the nuclear abundance of RelA and c-Rel (see Supplementary Fig. 5). We performed pulse-chase experiments using <sup>35</sup>S-labelled amino



**Figure 4** IKK $\alpha$  negatively regulates RelA and c-Rel nuclear accumulation and NF- $\kappa$ B target gene expression in LPS-stimulated macrophages. **a**, **b**, Total RNA was isolated from bone-marrow-derived macrophages at the indicated time points for real-time PCR analysis (**a**) and Rnase protection assays (**b**). **c**, Kinetics of NF- $\kappa$ B DNA-binding activity (top panel) and nuclear translocation of RelA and c-Rel proteins (bottom panel) in WT and

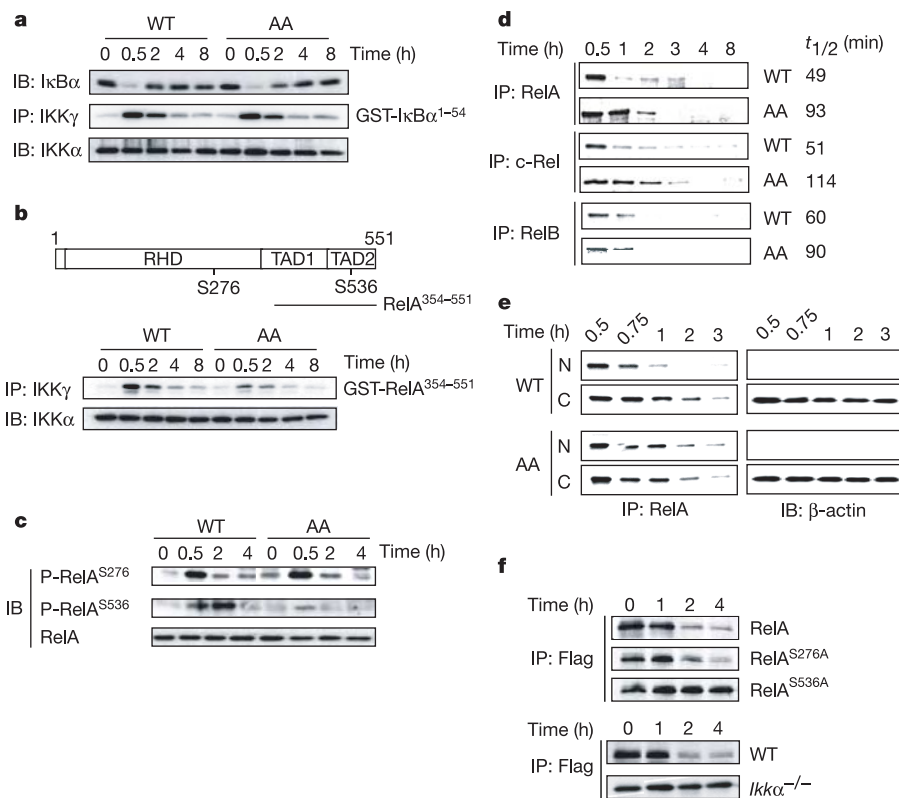
AA macrophages after LPS stimulation. **d**, Chromatin immunoprecipitation assays. Recruitment of RelA, c-Rel and RNA polymerase II to the Bfl-1, iNOS and IL-12p40 promoters was assessed by immunoprecipitation and PCR amplification of promoter sequences.

acids to track the turnover of Rel proteins in LPS-stimulated wild-type and *Ikk $\alpha$ <sup>AA/AA</sup>* macrophages. These experiments revealed a considerably increased half-life ( $t_{1/2}$ ) for both RelA and c-Rel in LPS-stimulated *Ikk $\alpha$ <sup>AA/AA</sup>* macrophages, whereas the turnover of RelB (an NF- $\kappa$ B subunit associated with anti-inflammatory activity<sup>27</sup>) was only modestly affected (Fig. 5d). Transfection studies revealed that replacing S536 of RelA with alanine abrogated LPS-induced turnover, but mutation of S276 had no effect (Fig. 5e). Furthermore, LPS-induced RelA turnover was abrogated in *Ikk $\alpha$ <sup>-/-</sup>* cells (Fig. 5e). Thus, IKK $\alpha$  activation accelerates RelA and c-Rel proteolysis and promotes their clearance from target gene promoters through their C-terminal phosphorylation. This mechanism may explain macrophage hyperactivity and increased inflammation in *Ikk $\alpha$ <sup>AA/AA</sup>* mice.

The studies described above reveal opposing yet complimentary roles for IKK $\alpha$  and IKK $\beta$  in the control of inflammation and innate immunity. We demonstrate a previously unknown function for IKK $\alpha$  in negative regulation of macrophage activation. This function might be mediated through IKK $\alpha$ -dependent phosphorylation of RelA and c-Rel, which results in accelerated turnover of these NF- $\kappa$ B subunits, thereby facilitating their removal from target gene promoters and terminating NF- $\kappa$ B-mediated gene induction. This model is supported by the protracted induction of a number of NF- $\kappa$ B target genes in *Ikk $\alpha$ <sup>AA/AA</sup>* macrophages and the extended retention time of RelA and c-Rel at their promoters. In addition, the half-life of RelA and c-Rel is significantly increased in *Ikk $\alpha$ <sup>AA/AA</sup>* macrophages, but the half-life of RelB associated with the termination of NF- $\kappa$ B-mediated gene induction<sup>28</sup>, is only modestly

affected. The connection between IKK $\alpha$ -dependent RelA C-terminal phosphorylation and stimulus-induced turnover is confirmed in transfection experiments that reveal S536 as a target for IKK $\alpha$ -dependent phosphorylation, showing that an S536A mutation abrogates LPS-induced RelA turnover. It is therefore evident that IKK $\alpha$ -dependent functions are associated with the termination of NF- $\kappa$ B-dependent transcription of pro-inflammatory genes. Impaired IKK $\alpha$  activation removes a 'brake' on macrophage activation and increases expression of pro-inflammatory genes, leading to elevated local inflammation, enhanced macrophage resistance to pathogen-induced apoptosis, and increased innate immunity to bacterial pathogens.

It would appear that the IKK complex has evolved to promote rapid but transient nuclear localization of Rel transcription factors in response to pro-inflammatory stimuli. Given that the first step in NF- $\kappa$ B activation, which depends on IKK $\beta$ , involves irreversible I $\kappa$ B $\alpha$  degradation<sup>3</sup>, a second step is necessary to ensure the transient nature of this signalling response. Part of this depends on inducible I $\kappa$ B $\alpha$  expression<sup>4,29</sup>, but that may not be sufficient. The IKK $\alpha$ -dependent step may have evolved to ensure the rapid turnover of pro-inflammatory RelA- and c-Rel-containing dimers and their replacement with anti-inflammatory RelB-containing dimers<sup>28</sup>. The cytoplasmic localization of the IKK complex would imply that C-terminal phosphorylation of RelA and c-Rel occurs in this compartment, while still associated with I $\kappa$ Bs; this might involve the newly discovered ELKS subunit<sup>30</sup>. However, the phosphorylated proteins are probably degraded mostly in the nucleus, as suggested by cell fractionation experiments (Fig. 4c). Selective inhibition of



**Figure 5** IKK $\alpha$  mediates C-terminal RelA phosphorylation and promotes its stimulus-induced turnover. **a**, **b**, Immune complex kinase assays performed with GST-I $\kappa$ B $\alpha$ (1–54) (**a**) and GST-RelA(354–551) (**b**). **c**, Immunoblot (IB) analysis of RelA phosphorylation. **d**, Pulse-chase analysis of RelA, c-Rel and RelB in LPS-stimulated WT and AA macrophages. Half-life ( $t_{1/2}$ ) was calculated from a semi-log plot of three independent experiments. **e**, Pulse-chase analysis of nuclear (N) and cytoplasmic (C) fractions of RelA

in LPS-stimulated WT and AA macrophages. **f**, Top panel, transient transfection with Flag-RelA, Flag-RelA(S276A) or Flag-RelA(S536A) mutants. Turnover of immunoprecipitated Flag-tagged proteins was assessed as above. Bottom panel, WT and *Ikk $\alpha$ <sup>-/-</sup>* murine embryonic fibroblast cells were transfected with Flag-RelA and protein turnover was assessed as above.



IKK $\alpha$  activation, although having no effect on I $\kappa$ B $\alpha$  degradation, has the unusual effect of enhancing innate immunity by preventing RelA and c-Rel turnover, thereby causing protracted NF- $\kappa$ B activation. This discovery may present new therapeutic opportunities for IKK $\alpha$  inhibitors in the treatment of complicated infections involving antibiotic resistance or compromised host immunity. □

## Methods

### Bacterial infections

The clinical GBS isolate, NCTC10/84 (serotype V)<sup>12</sup> was grown in Todd Hewitt broth (THB) without agitation at 37 °C to an absorbance at 600 nm of 0.4, equivalent to  $1 \times 10^8$  c.f.u. per ml. Bacteria collected by centrifugation were washed with sterile PBS. Mice were inoculated via the tail vein with  $5 \times 10^7$  c.f.u. GBS in 0.3 ml PBS. Blood was collected from mice after 24 h by retro-orbital bleed, serial dilutions were plated on Todd Hewitt agar (THA) in triplicate and colonies were counted. Mice were also inoculated intranasally with  $3 \times 10^7$  c.f.u. NCTC in 30  $\mu$ l PBS. These mice were killed at the indicated time points, and their tracheas were cannulated for bronchial-alveolar lavage with 3 aliquots of 0.8 ml ice-cold PBS. Serial dilutions of bronchial-alveolar lavage fluid were plated on THA in triplicate and c.f.u. counts determined. Lung tissue was also prepared for routine histological analysis.

### LPS-induced septic shock

Mice were challenged by intraperitoneal injection of 25 mg kg<sup>-1</sup> LPS (from *E. coli* serotype B5:055, Sigma) in pyrogen-free PBS. Chimaeric mice were generated using bone marrow from *Ikk $\alpha$ <sup>AA/AA</sup>* and wild-type littermate controls. Bone marrow cells were isolated in Hank's balanced salt solution (HBSS) and within 6 h,  $5 \times 10^6$  cells in 0.3 ml were injected into the tail vein of 8-week-old lethally irradiated wild-type hosts<sup>9</sup>. From 2 days before injection, host mice were housed under sterile conditions, using autoclaved cages, food, and water containing 25 mg l<sup>-1</sup> neomycin sulphate and 13 mg l<sup>-1</sup> polymyxin B sulphate.

### Macrophage isolation and stimulation

Bone marrow-derived macrophages (BMDMs) were generated as described<sup>17</sup>. Peritoneal macrophages were elicited by intraperitoneal injection of 3 ml of 3% thioglycollate in distilled water. After 3 days, cells were harvested and plated in RPMI 1640 medium supplemented with 10% heat-inactivated FCS, 100 U ml<sup>-1</sup> penicillin and streptomycin, and 2 mM glutamine. BMDMs were stimulated with 100 ng ml<sup>-1</sup> LPS for 30 min, after which the cells were washed and incubated in LPS-free media. For infection experiments, peritoneal macrophages were washed with antibiotic-free media and incubated with the indicated MOI (multiplicity of infection) of GBS, prepared as described above. Culture plates were centrifuged to bring bacteria into contact with macrophages and incubated for 2 h, after which cells were washed with media supplemented with 20  $\mu$ g ml<sup>-1</sup> gentamicin and incubated for a further 2 h. Macrophages were then washed and intracellular bacteria were released with 0.02% Triton X-100. Surviving bacteria were counted by plating serial dilutions in triplicate on THA. Apoptosis was measured by TUNEL assay using the ApoAlert kit (BD Bioscience). Nuclear morphology was assessed by DAPI counter-stain.

### Gene expression analysis and chromatin immunoprecipitation assays

Total cellular RNA was isolated using TRIzol (Invitrogen) and analysed by real-time PCR with SyBr Green (PE Biosystems 5700 thermocycler) or RNase protection assay. Primer sequences are available upon request. For real-time PCR analyses, all values were normalized to the level of cyclophilin mRNA. For RNase protection assays, total RNA was hybridized with RNA probes using a Riboquant Multiprobe RPA System (BD Bioscience), following manufacturer's instructions. ChIP assays were performed as described<sup>24</sup> using either anti-RelA (C-20), anti-c-Rel (C) or anti-Pol II (N-19) polyclonal antibodies (Santa Cruz) for immunoprecipitation. Sequences of promoter-specific primers and a detailed experimental protocol are available upon request.

### Kinase assay, immunoblotting and pulse-chase

Whole-cell lysates were prepared and IKK kinase activity was measured after immunoprecipitation with anti-IKK $\gamma$  (764) antibody (BD Pharmingen) as described<sup>9</sup>, using the following substrates: GST-I $\kappa$ B $\alpha$  (1–54), GST-RelA(354–551), GST-RelA(354–551; S536A), GST-c-Rel(422–588), GST-c-Rel(422–540). IKK recovery was determined by immunoblotting with anti-IKK $\alpha$  (M280) antibody (Santa Cruz). Immunoblotting was performed on gel-separated whole-cell lysates or on nuclear and cytoplasmic extracts<sup>9</sup>. For pulse-chase analyses, macrophages were labelled for 1 h with 100  $\mu$ Ci ml<sup>-1</sup> [<sup>35</sup>S]-methionine and [<sup>35</sup>S]-cysteine (ICN Biomedicals). The cells were washed and chased for the indicated time with fresh medium containing unlabelled amino acids. Cells were collected, cell lysates were pre-cleared using protein G-Sepharose, and proteins were immunoprecipitated with anti-RelA, anti-RelB, anti-cRel or anti-Flag antibodies. Immune complexes were resolved by SDS-PAGE and the gels were dried and autoradiographed. Radiolabelled protein bands were quantified using a phosphorimager.

### Transfection

Murine embryonic fibroblast (MEF) cells from wild-type or *Ikk $\alpha$ <sup>-/-</sup>* mice were cultured in 60-mm dishes and transfected with plasmid DNAs encoding Flag-RelA, Flag-RelA(S276A), or Flag-RelA(S536A) using LipofectAMINE Plus (Invitrogen, Gibco BRL) following the manufacturer's instructions.

Received 15 October 2004; accepted 17 February 2005; doi:10.1038/nature03491.

- Lawrence, T., Willoughby, D. A. & Gilroy, D. W. Anti-inflammatory lipid mediators and insights into the resolution of inflammation. *Nature Rev. Immunol.* **2**, 787–795 (2002).
- Bonizzi, G. & Karin, M. The two NF- $\kappa$ B activation pathways and their role in innate and adaptive immunity. *Trends Immunol.* **25**, 280–288 (2004).
- Karin, M. & Ben-Neriah, Y. Phosphorylation meets ubiquitination: the control of NF- $\kappa$ B activity. *Annu. Rev. Immunol.* **18**, 621–663 (2000).
- Hoffmann, A., Levchenko, A., Scott, M. L. & Baltimore, D. The I $\kappa$ B-NF- $\kappa$ B signaling module: temporal control and selective gene activation. *Science* **298**, 1241–1245 (2002).
- Wertz, I. E. *et al.* De-ubiquitination and ubiquitin ligase domains of A20 downregulate NF- $\kappa$ B signalling. *Nature* **430**, 694–699 (2004).
- Ghosh, S. & Karin, M. Missing pieces in the NF- $\kappa$ B puzzle. *Cell* **109** (suppl.), S81–S96 (2002).
- Makris, C. *et al.* Female mice heterozygous for IKK $\gamma$ /NEMO deficiencies develop a dermatopathy similar to the human X-linked disorder incontinentia pigmenti. *Mol. Cell* **5**, 969–979 (2000).
- Dejardin, E. *et al.* The lymphotoxin-beta receptor induces different patterns of gene expression via two NF- $\kappa$ B pathways. *Immunity* **17**, 525–535 (2002).
- Senfleben, U. *et al.* Activation by IKK $\alpha$  of a second, evolutionary conserved, NF- $\kappa$ B signaling pathway. *Science* **293**, 1495–1499 (2001).
- Anest, V. *et al.* A nucleosomal function for I $\kappa$ B kinase- $\alpha$  in NF- $\kappa$ B-dependent gene expression. *Nature* **423**, 659–663 (2003).
- Yamamoto, Y., Verma, U. N., Prajapati, S., Kwak, Y. T. & Gaynor, R. B. Histone H3 phosphorylation by IKK- $\alpha$  is critical for cytokine-induced gene expression. *Nature* **423**, 655–659 (2003).
- Liu, G. Y. *et al.* Sword and shield: Linked group B streptococcal  $\beta$ -hemolysin/cytolysin and carotenoid pigment function to subvert host phagocyte defense. *Proc. Natl Acad. Sci. USA* **101**, 14491–14496 (2004).
- Takeuchi, O. *et al.* Differential roles of TLR2 and TLR4 in recognition of gram-negative and gram-positive bacterial cell wall components. *Immunity* **11**, 443–451 (1999).
- Underhill, D. M. *et al.* The Toll-like receptor 2 is recruited to macrophage phagosomes and discriminates between pathogens. *Nature* **401**, 811–815 (1999).
- Bonizzi, G. *et al.* Activation of IKK $\alpha$  target genes depends on recognition of specific kappaB binding sites by RelB:p52 dimers. *EMBO J.* **23**, 4202–4210 (2004).
- Gordon, S. Pattern recognition receptors: doubling up for the innate immune response. *Cell* **111**, 927–930 (2002).
- Park, J. M., Greten, F. R., Li, Z. W. & Karin, M. Macrophage apoptosis by anthrax lethal factor through p38 MAP kinase inhibition. *Science* **297**, 2048–2051 (2002).
- Puliti, M., von Hunolstein, C., Bistoni, F., Orefici, G. & Tissi, L. Inhibition of nitric oxide synthase exacerbates group B streptococcus sepsis and arthritis in mice. *Infect. Immun.* **72**, 4891–4894 (2004).
- Sakurai, H. *et al.* Tumor necrosis factor- $\alpha$ -induced IKK phosphorylation of NF- $\kappa$ B p65 on serine 536 is mediated through the TRAF2, TRAF5, and TAK1 signaling pathway. *J. Biol. Chem.* **278**, 36916–36923 (2003).
- Martin, A. G. & Fresno, M. Tumor necrosis factor- $\alpha$  activation of NF- $\kappa$ B requires the phosphorylation of Ser-471 in the transactivation domain of c-Rel. *J. Biol. Chem.* **275**, 24383–24391 (2000).
- Fognani, C., Rondi, R., Romano, A. & Blasi, F. c-Rel-TD kinase: a serine/threonine kinase binding *in vivo* and *in vitro* c-Rel and phosphorylating its transactivation domain. *Oncogene* **19**, 2224–2232 (2000).
- Salghetti, S. E., Muratani, M., Wijnen, H., Futcher, B. & Tansey, W. P. Functional overlap of sequences that activate transcription and signal ubiquitin-mediated proteolysis. *Proc. Natl Acad. Sci. USA* **97**, 3118–3123 (2000).
- Perissi, V., Aggarwal, A., Glass, C. K., Rose, D. W. & Rosenfeld, M. G. A corepressor/coactivator exchange complex required for transcriptional activation by nuclear receptors and other regulated transcription factors. *Cell* **116**, 511–526 (2004).
- Saccani, S., Marazzi, I., Beg, A. A. & Natoli, G. Degradation of promoter-bound p65/RelA is essential for the prompt termination of the nuclear factor  $\kappa$ B response. *J. Exp. Med.* **200**, 107–113 (2004).
- Ryo, A. *et al.* Regulation of NF- $\kappa$ B signaling by Pin1-dependent prolyl isomerization and ubiquitin-mediated proteolysis of p65/RelA. *Mol. Cell* **12**, 1413–1426 (2003).
- Chen, E. *et al.* Degradation of proto-oncoprotein c-Rel by the ubiquitin-proteasome pathway. *J. Biol. Chem.* **273**, 35201–35207 (1998).
- Weih, F. *et al.* Multiorgan inflammation and hematopoietic abnormalities in mice with a targeted disruption of RelB, a member of the NF- $\kappa$ B/Rel family. *Cell* **80**, 331–340 (1995).
- Saccani, S., Pantano, S. & Natoli, G. Modulation of NF- $\kappa$ B activity by exchange of dimers. *Mol. Cell* **11**, 1563–1574 (2003).
- Sun, S. C., Ganchi, P. A., Ballard, D. W. & Greene, W. C. NF- $\kappa$ B controls expression of inhibitor I $\kappa$ B $\alpha$ : evidence for an inducible autoregulatory pathway. *Science* **259**, 1912–1915 (1993).
- Ducut Sigala, J. L. *et al.* Activation of transcription factor NF- $\kappa$ B requires ELKS, an I $\kappa$ B kinase regulatory subunit. *Science* **304**, 1963–1967 (2004).

**Supplementary Information** accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank H. Sakurai, H. Kamata, and G. Fresno for plasmid constructs, and J.M. Park and A. Hoffmann for discussions and advice. T.L. is the recipient of a Wellcome Trust International Research Fellowship. V.N. is an Edward J. Mallinckrodt, Jr Scholar. Research was supported by NIH grants to M.K., who is the Frank and Else Schilling American Cancer Society Research Professor.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to T.L. (t.lawrence@imperial.ac.uk).

# A unique clonal *JAK2* mutation leading to constitutive signalling causes polycythaemia vera

Chloé James<sup>1\*</sup>, Valérie Ugo<sup>1,2,3\*</sup>, Jean-Pierre Le Couédic<sup>1\*</sup>, Judith Staerk<sup>4</sup>, François Delhommeau<sup>1,3</sup>, Catherine Lacout<sup>1</sup>, Loïc Garçon<sup>1</sup>, Hana Raslova<sup>1</sup>, Roland Berger<sup>5</sup>, Annelise Bennaceur-Griscelli<sup>1,6</sup>, Jean Luc Villeval<sup>1</sup>, Stefan N. Constantinescu<sup>4</sup>, Nicole Casadevall<sup>1,3</sup> & William Vainchenker<sup>1,7</sup>

<sup>1</sup>INSERM U362, Institut Gustave Roussy, Paris XI University, PR1, 39 rue Camille Desmoulins, 94805 Villejuif Cedex, France

<sup>2</sup>Laboratoire d'Hématologie, CHU Brest, 29609 Brest Cedex, France

<sup>3</sup>Laboratoire d'Hématologie, Hôtel Dieu, AP-HP, 75181 Paris Cedex 04, France

<sup>4</sup>Ludwig Institute for Cancer Research and Christian de Duve Institute of Cellular Pathology & MEXP Unit, Université Catholique de Louvain, Brussels B-1200, Belgium

<sup>5</sup>INSERM E0210, Hôpital Necker, 75743 Paris Cedex 15, France

<sup>6</sup>Laboratoire d'Hématologie, Institut Gustave Roussy, 94805 Villejuif Cedex, France

<sup>7</sup>Polyclinique d'Hématologie, Hôpital Saint Louis, AP-HP, 75475 Paris Cedex 10, France

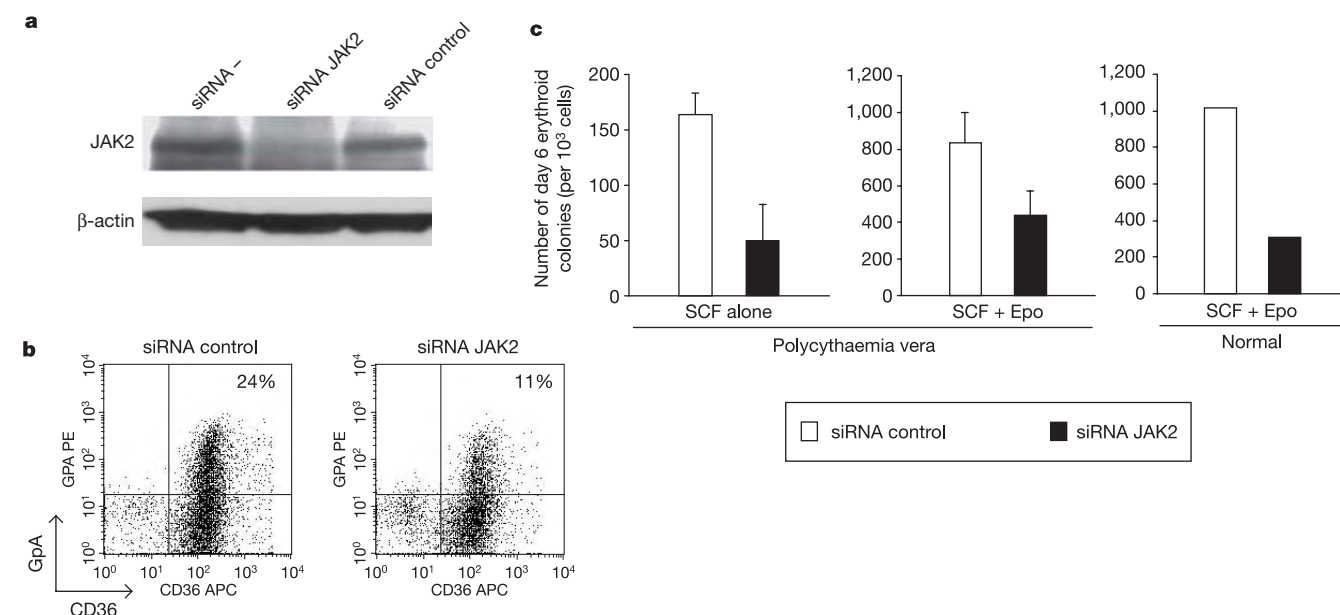
\* These authors contributed equally to this work

Myeloproliferative disorders are clonal haematopoietic stem cell malignancies characterized by independency or hypersensitivity of haematopoietic progenitors to numerous cytokines<sup>1,2</sup>. The molecular basis of most myeloproliferative disorders is unknown. On the basis of the model of chronic myeloid leukaemia, it is expected that a constitutive tyrosine kinase activity could be at the origin of these diseases. Polycythaemia vera is an acquired myeloproliferative disorder, characterized by the presence of polycythaemia diversely associated with thrombocytosis, leukocytosis and splenomegaly<sup>3</sup>. Polycythaemia vera progenitors are hypersensitive to erythropoietin and other cytokines<sup>4,5</sup>. Here, we describe a clonal and recurrent mutation in the

JH2 pseudo-kinase domain of the Janus kinase 2 (*JAK2*) gene in most (>80%) polycythaemia vera patients. The mutation, a valine-to-phenylalanine substitution at amino acid position 617, leads to constitutive tyrosine phosphorylation activity that promotes cytokine hypersensitivity and induces erythrocytosis in a mouse model. As this mutation is also found in other myeloproliferative disorders, this unique mutation will permit a new molecular classification of these disorders and novel therapeutic approaches.

In polycythaemia vera, the mechanisms leading to erythropoietin hypersensitivity and *in vitro* production of erythroid colonies in the absence of cytokines (referred to hereafter as EEC, for endogenous erythroid colonies)<sup>4</sup> are still unknown. We previously reported that inhibitors of *JAK2* (AG490), phosphatidylinositol-3-OH kinase (PI(3)K) and Src pathways hampered spontaneous erythroid terminal differentiation in polycythaemia vera<sup>6</sup>. We then focused on *JAK2*, an upstream molecule directly linked to erythropoietin receptor (EpoR) signalling<sup>7</sup>. In a first set of experiments, we used a short interfering RNA (siRNA) to knockdown *JAK2* expression. This siRNA, in contrast to a control siRNA, decreased *JAK2* protein levels to less than 10% in UT7 cells (Fig. 1a). It also impaired spontaneous erythroid differentiation in cells from polycythaemia vera patients, markedly inhibited EEC formation (Fig. 1b and c, left panel) and inhibited by 50% erythropoietin-dependent erythroid colony formation in polycythaemia vera cells (Fig. 1c, middle panel) and a normal sample (Fig. 1c, right panel).

This result suggests that *JAK2* has a principal role in EEC formation. Given these results, we searched for mutations in the *JAK2* gene. All the coding exons and intron–exon junctions were sequenced in three polycythaemia vera patients and two controls. In two of the polycythaemia vera patients, the analysis revealed one G-to-T mutation at nucleotide 1849 (in exon 12) leading to a substitution of valine to phenylalanine at position 617 (V617F; Fig. 2a), which is not a known polymorphism. In the third polycythaemia vera patient and the two controls no mutation was detected. To confirm this result, the 12th exon of *JAK2* was sequenced in 45 polycythaemia vera patients and 15 controls. Notably, the V617F substitution was found in 40 out of 45 poly-



**Figure 1** Erythropoietin-independent growth in polycythaemia vera cells is dependent on *JAK2*. **a**, Effects of a *JAK2* siRNA on *JAK2* protein levels in UT7 cells. **b**, The *JAK2* siRNA inhibits erythropoietin-independent acquisition of glycophorin A (GPA) in polycythaemia

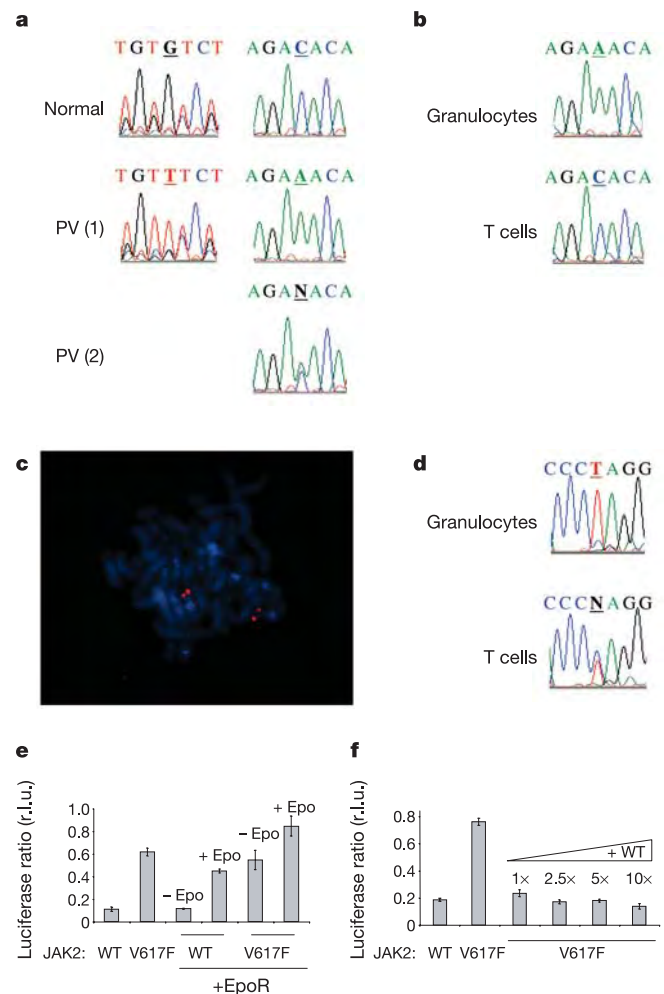
vera cells. **c**, The *JAK2* siRNA inhibits EEC formation and erythropoietin (Epo)-dependent erythroid colony formation in polycythaemia vera and normal CD36<sup>+</sup>/GpA<sup>+</sup> progenitors. Error bars indicate s.e.m.

cythaemia vera patients, but not in the 15 controls. In addition, this mutation was not found in samples from 35 patients with secondary erythrocytosis. To test whether the V617F mutation was acquired, we tested purified blood cell populations from three patients. The mutation was detected in myeloid lineages; that is granulocytes, erythroblasts differentiated from CD34<sup>+</sup> cells and platelets, but not in T cells (Fig. 2b). This result demonstrates that this mutation is acquired because it does not involve all cell types. The absence of detectable mutation in T cells does not preclude the possibility that a minority of T cells harbour this mutation. Alternatively, expression of the JAK2 mutant may be toxic for T cells, affecting the function and traffic of interferon (IFN)- $\gamma$  receptor 2 (ref. 8). In bone marrow cells from about 30% of the polycythaemia vera patients, only the mutated nucleotide was detected (Fig. 2a, PV(1)). This frequency is close to that reported for the chromosome 9 short arm (9p) loss of heterozygosity<sup>9</sup>. The polycythaemia vera clone appears to be largely predominant, but owing to the sensitivity of the technique, this does not exclude the presence of a minority of normal clones (<10%). However, our results strongly suggest that only the mutated gene is present in the polycythaemia vera clone. This could be due to a mutation associated to a deletion of the normal allele. To test this hypothesis, we performed a fluorescence *in situ* hybridization (FISH) analysis on myeloid cells from two patients without trisomy 9 and found that the *JAK2* gene was present on chromosome 9 of both patients (Fig. 2c). Thus, it was possible that the mutated gene was duplicated by a mitotic recombination, as suggested previously<sup>9</sup>. This hypothesis was supported by the fact that in these two patients with a polymorphism on exon 17 of *JAK2*, T cells displayed the two different alleles whereas granulocytes only expressed one (Fig. 2d). These results demonstrate that the mutated allele has been duplicated and the normal allele has been lost. Mitotic recombination may not be the only way to duplicate the mutated gene, because trisomy 9 is one of the most frequent cytogenetic abnormalities in polycythaemia vera sufferers<sup>10</sup>. Four patients with the V617F *JAK2* mutation and also with trisomy 9 were studied. In one of them, only the mutated allele was detected, suggesting a triplication of the gene. In the three others, the normal allele was detected at a level close to the mutated allele. This was also true in most (70%) polycythaemia vera bone marrow samples (Fig. 2a, PV(2)). As previously suggested in familial polycythemia<sup>11</sup>, this may indicate that development of polycythaemia vera requires at least two events, which may implicate the V617F mutation followed either by its duplication or by another uncharacterized genetic event. Alternatively, the detection of a normal *JAK2* allele may indicate the persistence of normal polyclonal haematopoiesis.

The V617F mutation is located in the JH2 pseudo-kinase domain of JAK2, which is involved in the auto-inhibition of its tyrosine kinase activity<sup>12</sup>. In this domain, three inhibitory regions have been described, one of which is located between amino acids 619 and 670 (ref. 13). Analysis of the predicted JAK2 structure indicates that amino acid 618 and surrounding residues interact and block the activation loop of the JH1 (kinase) domain<sup>14</sup>. Furthermore, Y570F and E695K mutations, both located in the JH2 domain, result in constitutive JAK2 activation<sup>15–17</sup>. To test whether the V617F mutation also enhances JAK2 activation, we studied whether the mutated JAK2 was able to spontaneously activate STAT transcription. JAK2- and STAT5-deficient gamma-2A cells<sup>18</sup> were co-transfected with complementary DNAs encoding a STAT5-dependent luciferase gene, human wild-type or V617F JAK2 and EpoR (Fig. 2e, f). The V617F JAK2 mutant was able to activate STAT-mediated transcription in the absence of erythropoietin (Fig. 2e), in contrast to wild-type JAK2. Co-transfection of EpoR did not significantly modify the ability of V617F JAK2 to activate STAT signalling molecules in the absence of ligand (Fig. 2e). Taken together, these results demonstrate that the V617F mutation

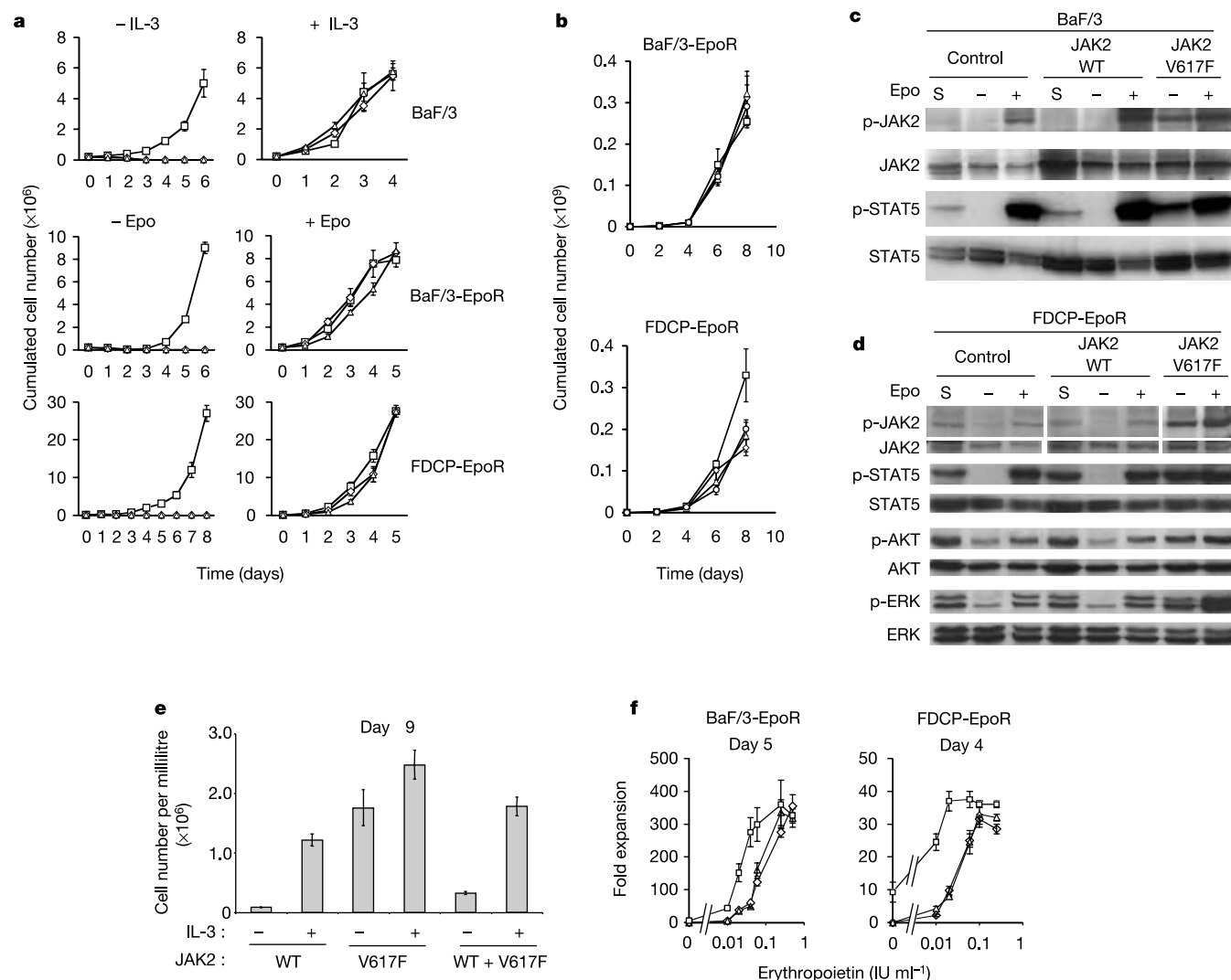
disrupts the auto-inhibitory activity of JAK2. In order to understand why the mutation is homozygous in at least 30% of polycythaemia vera patients, we hypothesized that wild-type JAK2 acts as a negative dominant of the V617F JAK2 mutant. Thus, we co-transfected both wild-type and mutated JAK2 cDNAs and observed that wild-type JAK2, at a 1:1 ratio to the mutant, abolished the constitutive activation of STAT5 induced by the mutant (Fig. 2f). These data show that loss of wild-type JAK2 confers a strong signalling advantage.

To demonstrate that the V617F mutation induces cytokine hypersensitivity, we expressed the JAK2 mutant in the murine BaF/3, BaF/3-EpoR and FDCP-EpoR factor-dependent cell lines. This induced growth factor independence after a short lag, whereas



**Figure 2** An acquired activating mutation in the 12th exon of JAK2. **a**, In polycythaemia vera patients, a G-to-T mutation at nucleotide 1849 of JAK2 leads to a V617F substitution. Patient PV(1) harbours only the mutated sequence, whereas patient PV(2) harbours both normal and mutated sequences. Sense and antisense sequences are shown. **b**, The JAK2 V617F substitution is found in granulocytes but not in T cells from a polycythaemia vera patient (antisense sequences). **c**, FISH analysis of the *JAK2* gene in one patient exhibiting only the mutated nucleotide. Two spots are detected in all mitosis. **d**, The exon 17 single nucleotide polymorphism (dbSNP 7048717) is found in T cells but not in granulocytes from one polycythaemia vera patient harbouring only the mutated nucleotide. **e**, STAT5 transcriptional activity (average of three replicates  $\pm$  s.d.) induced by the human V617F JAK2 mutant in JAK2-deficient gamma-2A cells is independent of cytokines and of the presence of EpoR. **f**, Wild-type JAK2 inhibits the constitutive STAT5 transcriptional activity (average of three replicates  $\pm$  s.d.) induced by the V617F JAK2 mutation in gamma-2A cells. r.l.u., relative light unit





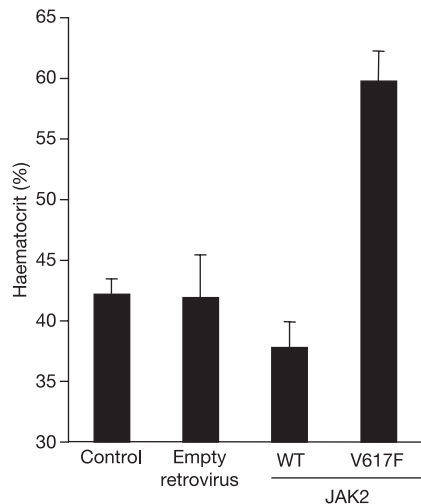
**Figure 3** Mutated JAK2 induces constitutive signalling leading to growth factor independence. **a**, Short-term growth of BaF/3, BaF/3-EpoR and FDCP-EpoR cells. Cells were sorted according to a similar GFP level after retroviral transduction and cultured in the presence or absence of cytokine. V617F JAK2 mutant, open squares; wild-type JAK2, open triangles; control retrovirus, open diamonds. **b**, Long-term growth of the V617F JAK2 BaF/3-EpoR and FDCP-EpoR cells selected for growth-factor independence. V617F JAK2 mutant cells without erythropoietin (open circles), with erythropoietin (open squares), wild-type JAK2 cells with erythropoietin (open triangles) and control retrovirus cells with erythropoietin (open diamonds) are shown. **c**, Studies of JAK2 and STAT5 phosphorylation by western blotting in wild-type BaF/3 cells, BaF/3 cells overexpressing wild-type JAK2 or mutated JAK2 cells (autonomous for growth). S, steady state. The

minus sign indicates 12 h after IL-3 deprivation, whereas the plus sign indicates after 100 ng ml<sup>-1</sup> IL-3 re-stimulation. **d**, Studies of JAK2, STAT5, ERK and AKT phosphorylation by western blotting in wild-type FDCP-EpoR cells, FDCP-EpoR cells overexpressing wild-type JAK2 or mutated JAK2 cells (autonomous for growth). The minus sign indicates 12 h after erythropoietin deprivation, whereas the plus sign indicates 10 IU ml<sup>-1</sup> erythropoietin re-stimulation. **e**, Co-expression of wild-type JAK2 inhibits the autonomous growth of BaF/3 cells (cell counts of three replicates  $\pm$  s.d.) induced by the V617F JAK2 mutant. **f**, Expression of the V617F JAK2 mutant (open squares) but not of wild-type JAK2 (open triangles) or an empty retrovirus (open diamonds) induces erythropoietin hypersensitivity in BaF/3-EpoR and FDCP-EpoR cells. Error bars in **a**, **b** and **f** indicate s.e.m.

control cells died within 36 h (Fig. 3a). These three cell lines could be maintained in culture for several weeks in the absence of growth factor with similar proliferation rates to those of wild-type cells stimulated by cytokines (Fig. 3b). This result could be reproduced in four independent experiments. Both human and murine mutated JAK2 induced factor-independent cell growth. In FDCP-EpoR and BaF/3 cells expressing the V617F JAK2 mutant and selected for cytokine-independent growth, we detected autophosphorylation of JAK2 associated with strong constitutive activation of STAT5 (Fig. 3c, d), but also activation of PI(3)K and ERK pathways in FDCP-EpoR cells (Fig. 3d). Interestingly, when wild-type JAK2 was transduced into BaF/3 cells rendered growth-factor independent by mutated JAK2, growth factor dependency was restored (Fig. 3e). This result further supports

the hypothesis of a competition between wild-type and mutated JAK2. As only a fraction of polycythaemia vera erythroid progenitors is associated with EEC formation, we asked whether the mutated JAK2 also confers erythropoietin hypersensitivity. The dose-response curve of FDCP-EpoR and BaF/3-EpoR cells to erythropoietin showed that mutated JAK2 induced an erythropoietin hypersensitivity with only a part of the cells being cytokine independent (Fig. 3f).

To understand the *in vivo* effects of this mutation, mice were transplanted with bone marrow cells infected with a murine embryonic stem cell virus (MESV)-derived retrovirus containing murine wild-type JAK2, murine V617F JAK2 mutant, or an empty vector, and were studied at 4 weeks after transplantation. Mice transplanted with the V617F mutant developed erythrocytosis



**Figure 4** Erythrocytosis induced in recipient mice after transplantation with bone marrow cells transduced with mutated JAK2. Haematocrit values (mean values  $\pm$  s.e.m.) were determined 4 weeks after bone marrow transplantation. More than 90% of the red blood cells were positive for GFP expression. Student's *t*-test:  $P = 0.003$  between the control ( $n = 5$ ) and the V617F JAK2 mice ( $n = 5$ );  $P = 0.0002$  between wild-type JAK2 ( $n = 4$ ) and V617F JAK2 mice.

(haematocrit: 60%) whereas those transplanted with wild-type JAK2 or the empty vector had a haematocrit level (40%) close to untransplanted mice (42%) (Fig. 4). This result underscores the role of this unique mutation in the pathogenesis of polycythaemia vera.

As polycythaemia vera, essential thrombocythaemia and idiopathic myelofibrosis are three closely related disorders<sup>1</sup>, we looked for the presence of the JAK2 V617F substitution in essential thrombocythaemia and idiopathic myelofibrosis patients. The mutation was found in 3 out of 7 idiopathic myelofibrosis and 9 of 21 essential thrombocythaemia patients studied, indicating that distinction between these three diseases is partly artificial. A search for other mutations in JAK2, other JAKs or other tyrosine kinases will be important in idiopathic myelofibrosis, essential thrombocythaemia and in polycythaemia vera patients negative for the V617F substitution.

What may then be the relationship between this JAK2 mutation and the pathogenesis of polycythaemia vera? Although it may not be the only defect in polycythaemia vera, this mutation has a principal role in the abnormal cytokine response and in the occurrence of polycythaemia. Indeed, the V617F JAK2 mutant is intrinsically active, contains all regions required for association with receptors and may become oligomerized to itself, as previously reported for the TEL-JAK2 fusion protein<sup>19</sup>. Furthermore, it could associate with the EpoR in the endoplasmic reticulum before receptor processing to the membrane, as was shown for wild-type JAK2 (ref. 20). Consequently, mutant JAK2 may induce EpoR activation in the cytoplasm, even in the absence of erythropoietin, but not at an optimum level. This may explain why the same mutation in polycythaemia vera may induce both a partial erythropoietin independence and cytokine hypersensitivity. In addition, as constitutively active JAK2 was shown to act cooperatively with a tyrosine kinase receptor in signalling to induce extensive self-renewal of multipotential haematopoietic cells<sup>21</sup>, the mutant JAK2 may also confer a proliferative advantage to the polycythaemia vera haematopoietic stem cells.

Surveying a large series of patients will precisely define the occurrence and role of this mutation in the different myeloproliferative disorders, whereas further development of animal

models will be useful in the development of new, targeted therapeutic approaches in these pathologies. □

## Methods

### Patient cells

The diagnosis of polycythaemia vera was based on the revised criteria of the Polycythemia Vera Study Group (PVSG)<sup>22</sup>. All patients exhibited a raised cell mass, an absence of cause of secondary erythrocytosis, a positive EEC assay and a splenomegaly, neutrophil leukocytosis, thrombocytosis, or a clonal marker. The diagnosis of the other myeloproliferative disorders was based on standard clinical criteria. Bone marrow and blood samples from patients and normal subjects were collected after informed consent was obtained. Purification and amplification of normal and polycythaemia vera progenitors were performed as described previously<sup>6</sup>.

### Western blotting

Western blot analysis was performed using conventional techniques with anti-JAK2 polyclonal antibodies (clone C-20, Santa Cruz) and with anti-phospho JAK2 Tyr 1007–1008, anti-phospho STAT5 Tyr 694, anti-phospho AKT Ser 473, anti-phospho ERK p42–p44 and anti-actin antibodies (all from Cell Signalling Technology).

### siRNA electroporation in polycythaemia vera progenitors

Day 5 CD36<sup>+</sup> polycythaemia vera cells grown in the presence of stem cell factor (SCF) and interleukin-3 (IL-3) were electroporated with siRNA targeted to JAK2 (Ambion; catalogue number 51118) or with a control siRNA (green fluorescent protein, GFP) by means of the nucleofactor technique (Amaxa Biosystems). After 24 h, CD36<sup>+</sup>/glycophorin A (GpA)<sup>−</sup> cells were plated in serum-free liquid medium or methylcellulose (StemCells Technologies) in the presence of 50 ng ml<sup>−1</sup> SCF, as described previously<sup>6</sup>.

### Separation of granulocytes, lymphocytes and platelets

Granulocytes, platelets and mononuclear cells were separated by standard techniques. CD3<sup>+</sup> mononuclear cells were separated by double-positive selection using a magnetic cell-sorting system (AutoMACS, Miltenyi Biotec).

### DNA sequencing

Genomic DNA was isolated from different cell fractions according to standard procedures and cDNAs were prepared from platelets. Each of the 23 exons of the *JAK2* gene (GenBank accession number AL161450) was amplified using standard polymerase chain reaction (PCR) conditions from 300 ng genomic DNA and primer sequences derived from flanking intronic sequences. PCR products were filtration purified (Multiscreen PCR, Millipore), sequenced using BigDye terminator cycle sequencing ready reaction kit (Applied Biosystems) according to the manufacturer's protocol and analysed on an ABI PRISM 3100 Genetic Analyser (see Supplementary Method 1).

### FISH analysis of JAK2

FISH analysis was performed both on metaphase and interphase nuclei with a human PAC *JAK2* probe (provided by P. Marynen)<sup>23</sup> containing exons 1–22 of the *JAK2* gene. Technical procedures were previously described<sup>24</sup>.

### Dual luciferase assays

Transcriptional activity of STAT5 was assessed by measuring luciferase expression of gamma-2A cells transfected with the pGRR5-Luciferase (pGRR5-Luc)<sup>25</sup>, which contains five copies of the STAT responsive elements of the IFN- $\gamma$  responsive region (GRR) of the high-affinity receptor for IgG1 promoter. Each experiment was performed in triplicate. For erythropoietin or mock stimulations, erythropoietin (100 international units (IU) per millilitre) was added 4 h after transfection (see Supplementary Method 2).

### Generation of JAK2 mutants

A human full-length *JAK2* open reading frame cloned in MSCV-GFP was obtained from J. Cools. Mutagenesis reactions for human and murine *JAK2* were performed using the QuickChange site-directed mutagenesis kit (Stratagene). All constructs were verified by sequencing.

### Cell lines and retroviral transductions

The gamma-2A human fibrosarcoma cells<sup>18</sup> were a gift from I. Kerr and G. Stark. Wild-type or mutant *JAK2* cDNAs were transfected into 293 EBNA or BOSC packaging cells to produce retroviruses<sup>26</sup>. After retroviral infection and cell sorting on GFP expression, cells were cultured in the presence or the absence of cytokines in RPMI medium complemented with 10% FCS. Cell numbers were recorded after Trypan blue dye exclusion staining.

### In vivo reconstitution of mouse haematopoietic system

Murine wild-type or V617F *JAK2* cDNAs were cloned in pMEGIX retroviral vector that contains, as a provirus, a MESV long terminal repeat driving the expression of the *JAK2* genes (cap-dependent) and the GFP reporter gene (encephalo-myocarditis virus- (ECMV)-derived IRES-dependent). We used the empty virus as a control. Viral particles were produced in the 293 EBNA cells. Bone marrow cells were collected from C57/B6 SJL mice (Charles River) 4 days after 5-fluorouracil treatment, infected for 5 days with the viral

particles and finally injected intravenously into lethally irradiated (950 rad) C57BL6 mice<sup>27</sup>. The haematocrit levels and the percentage of GFP-positive red blood cells were determined 4 weeks after transplantation.

Received 25 November 2004; accepted 14 March 2005; doi:10.1038/nature03546.

Published online 27 March 2005.

- Spivak, J. L. The chronic myeloproliferative disorders: clonality and clinical heterogeneity. *Semin. Hematol.* **41** (2 suppl. 3), 1–5 (2004).
- Prchal, J. T. Polycythemia vera and other primary polycythemas. *Curr. Opin. Hematol.* **12**, 112–116 (2005).
- Spivak, J. L. Polycythemia vera: myths, mechanisms, and management. *Blood* **100**, 4272–4290 (2002).
- Prchal, J. F. & Axelrad, A. A. Bone-marrow responses in polycythemia vera. *N. Engl. J. Med.* **290**, 1382 (1974).
- Casadevall, N. *et al.* Erythroid progenitors in polycythemia vera. Demonstration of their hypersensitivity to erythropoietin using serum-free cultures. *Blood* **59**, 447–451 (1982).
- Ugo, V. *et al.* Multiple signaling pathways are involved in erythropoietin-independent differentiation of erythroid progenitors in polycythemia vera. *Exp. Hematol.* **32**, 179–187 (2004).
- Wittkuhn, B. A. *et al.* JAK2 associates with the erythropoietin receptor and is tyrosine phosphorylated and activated following stimulation with erythropoietin. *Cell* **74**, 227–236 (1993).
- Bernabei, P. *et al.* Interferon-gamma receptor 2 expression as the deciding factor in human T, B, and myeloid cell proliferation or death. *J. Leukoc. Biol.* **70**, 950–960 (2001).
- Kralovics, R., Guan, Y. & Prchal, J. T. Acquired uniparental disomy of chromosome 9p is a frequent stem cell defect in polycythemia vera. *Exp. Hematol.* **30**, 229–236 (2002).
- Najfeld, V., Montella, L., Scalise, A. & Fruchtman, S. Exploring polycythemia vera with fluorescence in situ hybridization: additional cryptic 9p is the most frequent abnormality detected. *Br. J. Haematol.* **119**, 558–566 (2002).
- Kralovics, R., Stockton, D. W. & Prchal, J. T. Clonal hematopoiesis in familial polycythemia vera suggests the involvement of multiple mutational events in the early pathogenesis of the disease. *Blood* **102**, 3793–3796 (2003).
- Saharinen, P., Takaluoma, K. & Silvennoinen, O. Regulation of the Jak2 tyrosine kinase by its pseudokinase domain. *Mol. Cell. Biol.* **20**, 3387–3395 (2000).
- Saharinen, P., Vihinen, M. & Silvennoinen, O. Autoinhibition of Jak2 tyrosine kinase is dependent on specific regions in its pseudokinase domain. *Mol. Biol. Cell* **14**, 1448–1459 (2003).
- Lindauer, K., Loerting, T., Liedl, K. R. & Kroemer, R. T. Prediction of the structure of human Janus kinase 2 (JAK2) comprising the two carboxy-terminal domains reveals a mechanism for autoregulation. *Protein Eng.* **14**, 27–37 (2001).
- Argesinger, L. S. *et al.* Autophosphorylation of JAK2 on tyrosines 221 and 570 regulates its activity. *Mol. Cell. Biol.* **24**, 4955–4967 (2004).
- Feener, E. P., Rosario, F., Dunn, S. L., Stancheva, Z. & Myers, M. G. J. Tyrosine phosphorylation of Jak2 in the JH2 domain inhibits cytokine signaling. *Mol. Cell. Biol.* **24**, 4968–4978 (2004).
- Luo, H. *et al.* Mutation in the Jak kinase JH2 domain hyperactivates *Drosophila* and mammalian Jak-Stat pathways. *Mol. Cell. Biol.* **17**, 1562–1571 (1997).
- Kohlhuber, F. *et al.* A JAK1/JAK2 chimera can sustain alpha and gamma interferon responses. *Mol. Cell. Biol.* **17**, 695–706 (1997).
- Lacronique, V. *et al.* A TEL-JAK2 fusion protein with constitutive kinase activity in human leukemia. *Science* **278**, 1309–1312 (1997).
- Huang, L. J., Constantinescu, S. N. & Lodish, H. F. The N-terminal domain of Janus kinase 2 is required for Golgi processing and cell surface expression of erythropoietin receptor. *Mol. Cell* **8**, 1327–1338 (2001).
- Zhao, S. *et al.* JAK2, complemented by a second signal from c-kit or flt-3, triggers extensive self-renewal of primary multipotential hemopoietic cells. *EMBO J.* **21**, 2159–2167 (2002).
- Pearson, T. C. & Messiney, M. The diagnostic criteria of polycythemia rubra vera. *Leuk. Lymphoma* **22** (suppl. 1), 87–93 (1996).
- Cools, J. *et al.* Genomic organization of human JAK2 and mutation analysis of its JH2-domain in leukemia. *Cytogenet. Cell Genet.* **85**, 260–266 (1999).
- Le Coniat, M., Romana, S. P. & Berger, R. Partial chromosome 21 amplification in a child with acute lymphoblastic leukemia. *Genes Chromosom. Cancer* **14**, 204–209 (1995).
- Dumoutier, L., Van Roost, E., Colau, D. & Renauld, J. C. Human interleukin-10-related T cell-derived inducible factor: molecular cloning and functional characterization as an hepatocyte-stimulating factor. *Proc. Natl Acad. Sci. USA* **97**, 10144–10149 (2000).
- Chagraoui, H. *et al.* Stimulation of osteoprotegerin production is responsible for osteoclasts in mice overexpressing TPO. *Blood* **101**, 2983–2989 (2003).
- Dorsch, M. *et al.* Ectopic expression of delta4 impairs hematopoietic development and leads to lymphoproliferative disease. *Blood* **100**, 2046–2055 (2002).

**Supplementary Information** accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** The authors are grateful to M.-H. Courtier, E. Leclerc and A. Tonon for technical assistance, P. Marynen and J. Cools for providing the human JAK2 cDNA, and J. Feunteun, F. Wendling and O. Bernard for scientific discussions. We thank I. Teysandier and C. Marzac for their help in collecting polycythemia vera samples, and J.-C. Brouet, S. Cheze, J.-J. Kiladjian, F. Lellouche, M. Leporrier, M. Macro, P. Morel, O. Reman, L. Roy, A.-L. Taksin, B. Varet and J.-P. Vilque for their help in collecting samples and clinical data. We are also grateful to the patients for their agreement in participating in this study. This work was supported by grants from La Ligue Nationale contre le Cancer (équipe labellisée 2003), la Fédération belge contre le cancer and the FNRS, Belgium. C.J. was supported by a fellowship from the Fondation pour la Recherche Médicale. J.S. was a recipient of a Marie Curie fellowship and of a Daimler-Benz PhD fellowship. S.N.C. is a Research Associate of the FNRS. W.V. is supported by an interface contract between INSERM and IGR.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to W.V. ([verpre@igr.fr](mailto:verpre@igr.fr)).

## Peak SIV replication in resting memory CD4<sup>+</sup> T cells depletes gut lamina propria CD4<sup>+</sup> T cells

Qingsheng Li<sup>1</sup>, Lijie Duan<sup>1</sup>, Jacob D. Estes<sup>1</sup>, Zhong-Min Ma<sup>4</sup>, Tracy Rourke<sup>4</sup>, Yichuan Wang<sup>4</sup>, Cavan Reilly<sup>2</sup>, John Carlis<sup>3</sup>, Christopher J. Miller<sup>4,5</sup> & Ashley T. Haase<sup>1</sup>

<sup>1</sup>Department of Microbiology, Medical School, University of Minnesota, MMC 196, 420 Delaware Street S.E., <sup>2</sup>Division of Biostatistics, School of Public Health, University of Minnesota, MMC 303, 420 Delaware Street S.E., and <sup>3</sup>Department of Computer Science and Engineering, Institute of Technology, University of Minnesota, 200 Union Street S.E., Minneapolis, Minnesota 55455, USA

<sup>4</sup>California National Primate Research Center and Center for Comparative Medicine, and <sup>5</sup>Department of Pathology, Microbiology and Immunology, School of Veterinary Medicine, and Division of Infectious Diseases, School of Medicine, University of California, Davis, California 95616, USA

In early simian immunodeficiency virus (SIV) and human immunodeficiency virus-1 (HIV-1) infections, gut-associated lymphatic tissue (GALT), the largest component of the lymphoid organ system<sup>1</sup>, is a principal site of both virus production and depletion of primarily lamina propria memory CD4<sup>+</sup> T cells; that is, CD4-expressing T cells that previously encountered antigens and microbes and homed to the lamina propria of GALT<sup>2–9</sup>. Here, we show that peak virus production in gut tissues of SIV-infected rhesus macaques coincides with peak numbers of infected memory CD4<sup>+</sup> T cells. Surprisingly, most of the initially infected memory cells were not, as expected<sup>10,11</sup>, activated but were instead immunophenotypically ‘resting’ cells that, unlike truly resting cells, but like the first cells mainly infected at other mucosal sites and peripheral lymph nodes<sup>12,13</sup>, are capable of supporting virus production. In addition to inducing immune activation and thereby providing activated CD4<sup>+</sup> T-cell targets to sustain infection, virus production also triggered<sup>14</sup> an immunopathologically limiting Fas–Fas-ligand-mediated apoptotic pathway<sup>15,16</sup> in lamina propria CD4<sup>+</sup> T cells, resulting in their preferential ablation. Thus, SIV exploits a large, resident population of resting memory CD4<sup>+</sup> T cells in GALT to produce peak levels of virus that directly (through lytic infection) and indirectly (through apoptosis of infected and uninfected cells) deplete CD4<sup>+</sup> T cells in the effector arm of GALT. The scale of this CD4<sup>+</sup> T-cell depletion has adverse effects on the immune system of the host, underscoring the importance of developing countermeasures to SIV that are effective before infection of GALT.

We investigated virus production and mechanisms of CD4<sup>+</sup> T-cell depletion in GALT of rhesus macaques infected intravaginally with the SIVmac251 virus. We focused our analysis on the colon as representative of GALT with organized inductive sites in scattered follicular aggregates and effector sites in lamina propria. As described in detail elsewhere<sup>26</sup>, virus production, measured as copies of SIV RNA per microgram of tissue RNA, peaked at day 10 after inoculation, and then declined about 20-fold by day 28 after inoculation, the last time point examined. This viral peak coincided with the peak in SIV RNA-positive cells (Fig. 1), in both the follicular inductive and diffuse effector arms of GALT (Fig. 2). Although there were SIV RNA-positive cells at both sites, massive depletion of CD4<sup>+</sup> T cells was confined to lamina propria (Fig. 2). Depletion of CD4<sup>+</sup> T cells in lamina propria, already detectable at day 6 after inoculation, continued at a rapid rate between days 8 and 14 after inoculation and reached a plateau that was about 70% below baseline levels (Fig. 1), corresponding to the selective loss of essentially the entire lamina propria CD45RO<sup>+</sup> memory CD4<sup>+</sup> T-cell population (Fig. 3).



One hypothesis originally advanced to account for this early, preferential depletion of memory  $CD4^+$  T cells in GALT<sup>2,10,11</sup> is that a large proportion of this population is activated in GALT, because of constant antigenic stimulation, and thus presents the virus with a favoured target cell for replication.  $CD4^+$  T-cell depletion is then thought to be the result of infection and subsequent killing of this large pool of susceptible hosts. We tested this hypothesis by characterizing the types and activation state of productively infected cells. About 90% of the SIV RNA-positive cells in both lamina propria and follicular aggregates were  $CD4^+$  T cells (data not shown), of which 95% were  $CD45RO^+$  memory cells (Fig. 3a). In the follicular aggregates, in which there are both naive  $CD45RA^+$  and memory  $CD4^+$  T cells, all of the SIV RNA-positive cells were also  $CD45RO^+$  memory cells (data not shown).

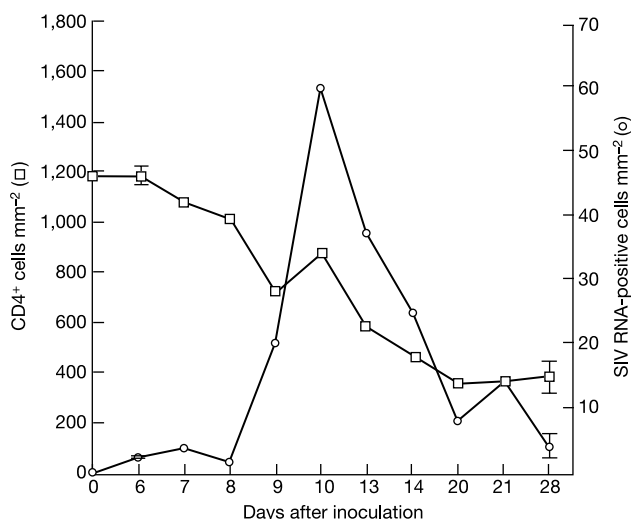
Although most of the SIV RNA-positive cells were memory  $CD4^+$  T cells, direct typing of the infected cells in tissue sections revealed that they were not mainly activated and proliferating, contrary to previous proposals<sup>2,10,11</sup>. We characterized the SIV RNA-positive cells with markers of T-cell activation and proliferation used in other studies of target cells in primary SIV infection<sup>11–13</sup>:  $CD69$ , a marker of early activation displaying transient expression<sup>17,18</sup>, and  $CD25$  and  $Ki67$ , markers for which expression is sustained in activated and proliferating cells. At peak viral production (day 10 after inoculation), 91–93% of the SIV RNA-positive memory  $CD4^+$  T cells in colon displayed the  $CD69^- CD25^- Ki67^-$  phenotype (Fig. 3b–d).

Thus, most of the first cells productively infected in GALT are memory  $CD4^+$  T cells; these cells are neither activated nor proliferating according to available markers, similar to most cells initially infected in cervico-vaginal mucosa and peripheral lymph nodes<sup>12,13</sup>. We refer to these cells as resting, because we believe that they are a recently activated population retaining—as they return to a resting state—sufficient levels of CCR5 co-receptor, nucleotide pools and transcriptional activators to enable them, unlike truly quiescent  $CD4^+$  T cells, to support productive infection<sup>13</sup>. Although they have about fivefold lower levels of SIV RNA per cell<sup>12</sup> and cell-

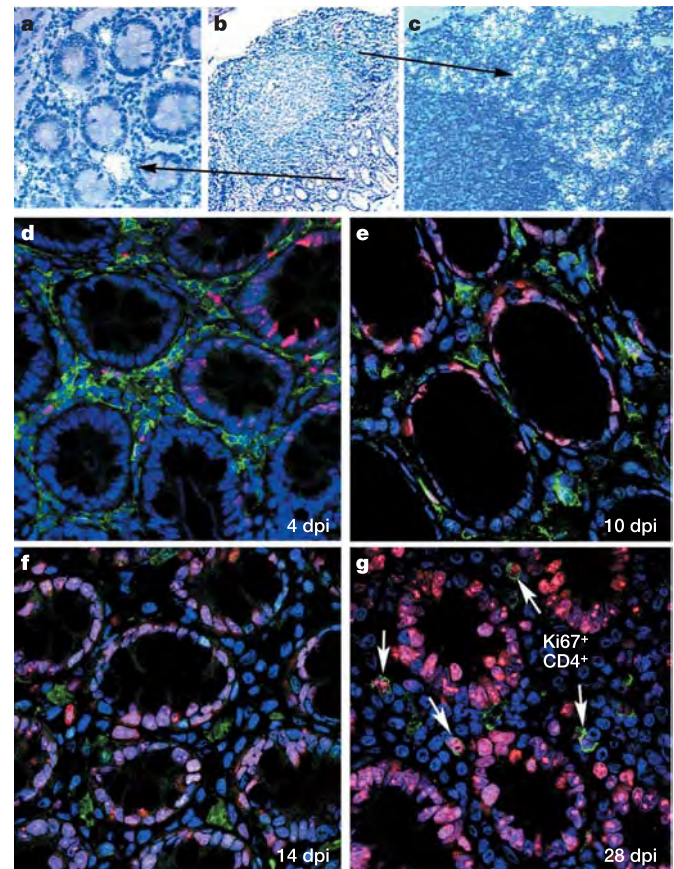
associated viral particles<sup>13</sup> compared with infected, activated  $CD4^+$  T cells in tissue section ‘snapshots’, these infected resting cells outnumber the infected activated cells by an order of magnitude and are the major contributor to the peak virus production of early infection (see below).

The peak in SIV RNA at day 10 after inoculation coincided with the peak number ( $55 \text{ cells mm}^{-2}$ ) and proportion (93%) of these resting SIV RNA-positive  $Ki67^- CD4^+$  T cells (Fig. 3e, g) and with the initial predominance of uninfected  $Ki67^-$  over  $Ki67^+$   $CD4^+$  T cells (Fig. 3f). The resulting infected cell population then decayed with a half-life of 4 days, to less than  $1 \text{ cell mm}^{-2}$  at day 28 after inoculation (Fig. 3e), in parallel with the contraction in size of the uninfected  $Ki67^- CD4^+$  T-cell population (Fig. 3f). The proportion of SIV RNA-positive activated and proliferating  $Ki67^+ CD4^+$  T cells reciprocally increased, from 10% at day 10 to 90% at day 28 after inoculation (Fig. 3g), but at peak was only one-tenth the size of the infected resting population. Moreover, this infected, activated population did not increase substantially despite the increased proportion of  $Ki67^+ CD4^+$  T cells (illustrated in Fig. 2); this was partly due to depletion of, and thus an overall decrease in, the number of uninfected  $Ki67^+ CD4^+$  T-cell targets in the colon (Fig. 3f).

These shifts in infected and uninfected populations, and post-



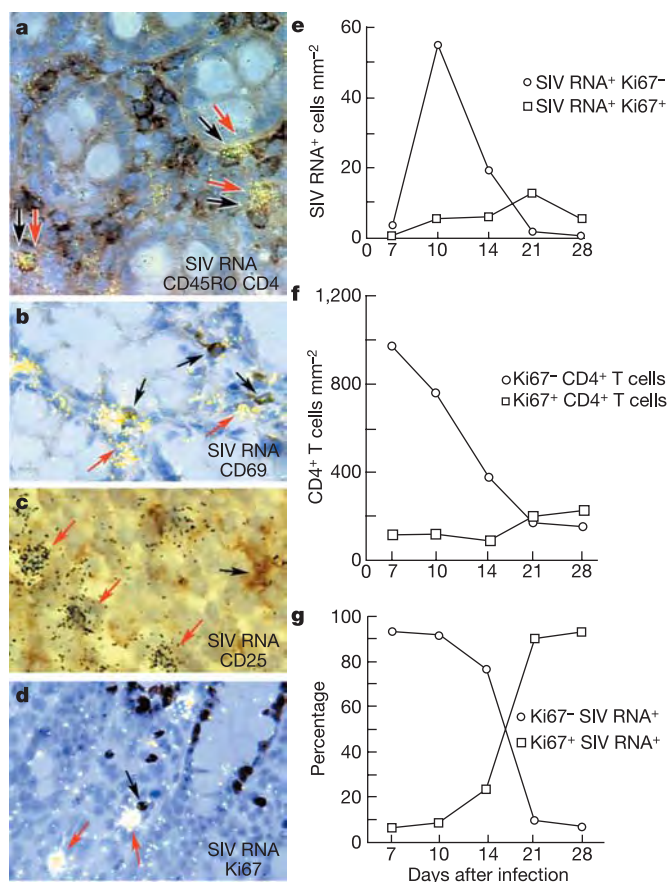
**Figure 1** SIV RNA-positive productively infected cells and  $CD4^+$  T-cell depletion in colon lamina propria. SIV RNA-positive cells were detected by *in situ* hybridization and enumerated in measured areas of sections of colon obtained at necropsy at the times shown. The number of  $CD4^+$  T cells per  $\text{mm}^2$  was determined by counting immunofluorescently stained cells in confocal images in defined areas, illustrated in Fig. 2. Error bars ( $\pm 1$  s.d.) are shown for days 6 and 28 for the two and three animals, respectively, necropsied at these times.



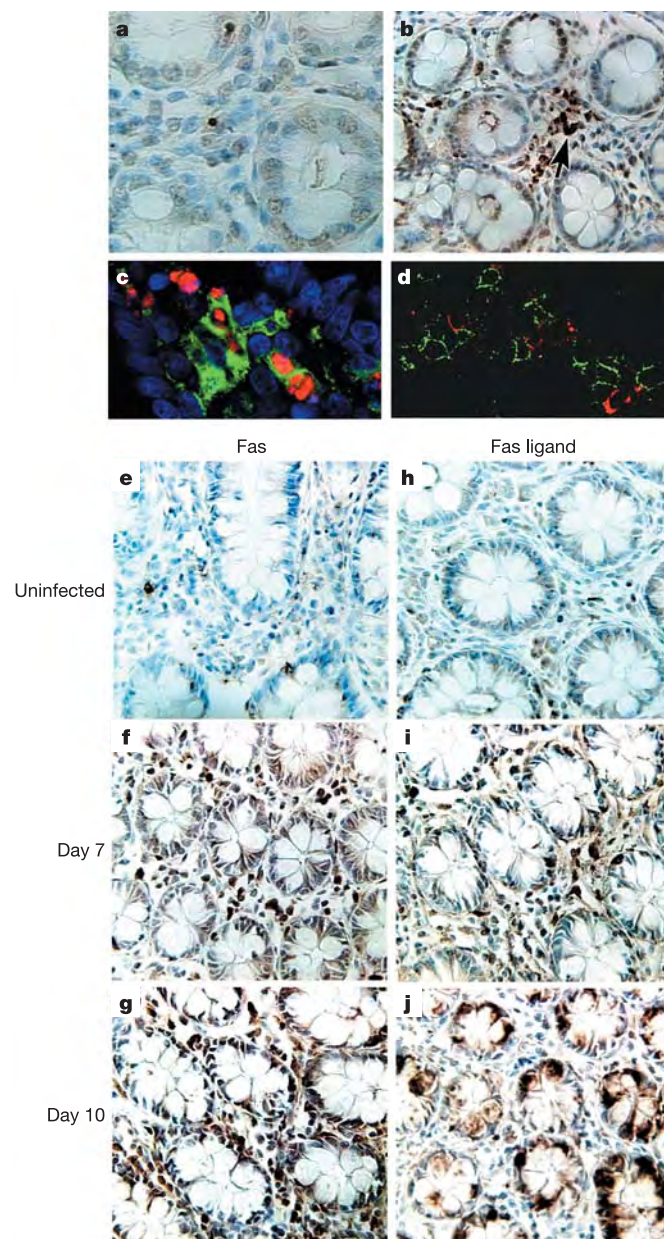
**Figure 2** Productive infection in organized and diffuse lamina propria, and  $CD4^+$  T-cell depletion in lamina propria. **a–c**, Arrows in the low-magnification image (**b**) point to white-appearing (from epipolarized light reflected from silver grains in radioautographs) SIV RNA-positive cells detected by *in situ* hybridization in lamina propria (**a**), and to infected cells in the T-cell area of a follicular aggregate (**c**). **d–g**, Immunofluorescent images of  $CD4^+$  T cells (green) and  $Ki67^+$  cells (red). Cell nuclei are counterstained blue. Rapid depletion of  $CD4^+$  cells is evident by days 10–14 post infection (dpi) (**e, f**).  $Ki67^+ CD4^+$  T cells (arrows) are increased at day 28 (**g**). Original magnification: **a**,  $\times 200$ ; **b**,  $\times 40$ ; **c**,  $\times 100$ ; **d–g**,  $\times 600$ .

peak declines in viral RNA levels and numbers of infected cells, are in accord with the predictions of models of substrate availability and exhaustion<sup>12,13,19</sup>. We revise these models here to reflect a new characterization of the exhausted substrate as resting rather than activated CD4<sup>+</sup> T cells, the loss of which was originally proposed to be the mechanism responsible for reduction in plasma viraemia in the acute stage of HIV-1 infection<sup>19</sup>. In the substrate availability model<sup>12,13</sup>, relatively large numbers of ostensibly resting memory CD4<sup>+</sup> T cells are preferentially infected in early infection because they constitute the largest population of susceptible target cells available for propagation of infection. We have hitherto proposed that the main role of these resting cells is as 'placeholders', maintaining unbroken chains of transmission until immune activation supplies activated CD4<sup>+</sup> T-cell targets enabling SIV and HIV-1 to replicate to higher levels<sup>12,13</sup>. We now show that infected resting T cells actually contribute substantially to peak virus production because of the relatively large number of these cells that are initially

available and infected. Substrate exhaustion of this pool of resting T-cell targets, which is capable of supporting replication, is then responsible for the 20-fold decrease in SIV RNA levels<sup>26</sup> from the peak (day 10 after inoculation) to last point examined (day 28 after inoculation)—the infected, resting CD4<sup>+</sup> T cells are eliminated and not replaced owing to the state of heightened immune activation that ensues as a result of early virus production. This provides a larger proportion of activated CD4<sup>+</sup> T cells, and results in a higher proportion of activated CD4<sup>+</sup> T cells that expresses SIV RNA,



**Figure 3** Initial infection of memory CD4<sup>+</sup> T cells, subsequent infection of activated cells, substrate availability and exhaustion. **a–d**, SIV replicates mainly in resting memory CD4<sup>+</sup> T cells at peak production. Combined *in situ* hybridization/immunohistochemical staining. Red arrows indicate SIV RNA-positive cells overlaid by silver grains (yellow/white in epipolarized light, black in transmitted light). Black arrows indicate brown-stained CD45RO<sup>+</sup> memory CD4<sup>+</sup> T cells (**a**) and CD69<sup>+</sup>, CD25<sup>+</sup> and Ki67<sup>+</sup> cells (**b–d**). **a**, Double-labelled SIV RNA-positive CD45RO<sup>+</sup> memory CD4<sup>+</sup> T cells. **b–d**, Singly labelled SIV RNA-positive CD69<sup>+</sup> CD25<sup>+</sup> Ki67<sup>+</sup> cells. Original magnification: **a**, **c**,  $\times 400$ ; **b**, **d**,  $\times 200$ . **e–g**, Substrate availability. Reciprocal changes in SIV RNA-positive Ki67<sup>+</sup>/Ki67<sup>-</sup> populations and shift from predominantly Ki67<sup>-</sup> SIV RNA-positive cells at peak production to predominantly Ki67<sup>+</sup> cells at day 28 (**e**, **g**). The increase in the number of SIV-positive/negative Ki67<sup>+</sup> cells (**e**, **f**) is limited by CD4<sup>+</sup> T-cell depletion. The number and percentage of SIV-positive/negative Ki67<sup>+</sup>/Ki67<sup>-</sup> cells per mm<sup>2</sup> were determined as described in the Methods.



**Figure 4** CD4<sup>+</sup> T-cell depletion in lamina propria by Fas–Fas-ligand-mediated apoptosis. **a–d**, Apoptosis of lamina propria CD4<sup>+</sup> T cells. Increased frequency of TUNEL-positive cells (indicated by an arrow) in lamina propria from 1 day after inoculation (**a**) to 10 days (**b**). **c**, Apoptotic activated caspase-3-positive cells (red) are CD4<sup>+</sup> T cells (green). **d**, CD8<sup>+</sup> T cells (green) are activated caspase-3-negative cells. **e–j**, Increases in Fas and Fas ligand in T cells. Compared with uninfected animals (**e**, **h**) and earlier time points, the frequency of brown-stained Fas-positive (**f**, **g**) and Fas-ligand-positive (**i**, **j**) lamina propria T cells increases markedly between days 7 and 10 after inoculation. Later increases in Fas ligand in colonic epithelial cells (**j**) are related to enteropathic changes (A.T.H., manuscript in preparation). Original magnification: **a**, **b**, **e–j**,  $\times 400$ ; **c**, **d**,  $\times 600$ .



corresponding to the increased availability of these cellular targets (Fig. 3f, g).

Although infection of activated CD4<sup>+</sup> T cells provides a mechanism to sustain virus production, the levels of viral RNA and frequency of infected cells fall despite higher levels of replication in activated cells; for example, in snapshots of infection *in vivo*, infected activated cells are enveloped by five times as many virions compared to resting cells<sup>13</sup>. Thus, virus production might be expected to increase. One reason that this is not the case is that CD4<sup>+</sup> T-cell depletion markedly reduces the size of the population of susceptible cells available for infection, and also the spatial proximity of infected to susceptible cells. The latter factor is probably important in efficiently propagating infection *in vivo* inasmuch as infected cells at peak production are clustered to within one to two cell diameters of one another<sup>20</sup>. Although elimination of infected cells by virus-specific cytotoxic T lymphocytes responding to immunodominant epitopes in Gag and Tat (which constitute about 70% of the cytotoxic T lymphocyte response in acute SIV infection of Mamu A\*01 animals (ref. 21)) is another possible reason for the decrease in SIV RNA production, it is unlikely because, as we show elsewhere, responses to these epitopes are minimal or undetectable in colon<sup>22</sup>. Additional studies will be needed to determine whether the cause of decreased virus production despite replication in activated cells is related to cytotoxic T lymphocyte responses to other epitopes, cytokine-mediated suppression of viral replication, or other mechanisms.

By what mechanisms are CD4<sup>+</sup> T cells depleted in the lamina propria of GALT? Death via cytopathic effects of viral replication in productively infected cells certainly contributes to depletion, but this mechanism alone cannot account for the preferential depletion of lamina propria versus follicular GALT or the magnitude of the depletion in lamina propria. At peak levels, only 7% of the CD4<sup>+</sup> T cells were SIV RNA-positive, and the total number of both resting and activated cells infected and lost from days 6–28 after inoculation could have accounted for at most 20% of the depletion of ~800 CD4<sup>+</sup> T cells mm<sup>-2</sup>. Instead, we suspected, for the following two reasons, that large numbers of lamina propria CD4<sup>+</sup> T cells were being selectively eliminated by Fas–Fas-ligand-mediated apoptosis of this population: (1) lamina propria T cells are particularly prone to Fas-mediated apoptosis as a normal immunoregulatory mechanism to control activation and inflammation in the gastrointestinal tract<sup>15,16</sup>; and (2) because lamina propria CD4<sup>+</sup> T cells selectively upregulate Fas and Fas ligand when exposed *in vitro* to recombinant envelope gp120 (ref. 14), enhanced expression *in vivo* of Fas and Fas ligand would be expected in large numbers of lamina propria CD4<sup>+</sup> T cells exposed to virion gp120 during peak virus production.

We sought evidence for the suspected selective apoptosis of CD4<sup>+</sup> T cells in lamina propria compared with follicular GALT by monitoring changes in the proportion of TdT-mediated dUTP nick end labelling (TUNEL)-positive and activated caspase-3-positive T cells. We documented marked increases from baseline levels in CD4<sup>+</sup>, but not CD8<sup>+</sup>, T cells undergoing apoptosis in lamina propria (Fig. 4a–d), which paralleled the depletion of CD4<sup>+</sup> T cells there, but only minimal increases in apoptosis in follicular GALT, where CD4<sup>+</sup> T-cell depletion was insignificant (data not shown). We also documented increased expression of the mediators of T-cell apoptosis (Fas and Fas ligand) in lamina propria; this expression paralleled the time course and increase in number of apoptotic cells (Fig. 4b, e–j) as well the rapid decrease in the number of CD4<sup>+</sup> T cells between 8 and 14 days after inoculation (Fig. 2d–g). Again, these changes were selective for lamina propria.

Because SIV and HIV-1 can productively infect recently activated but ostensibly resting memory CD4<sup>+</sup> T cells, they can take immediate advantage of the availability of a large population of susceptible

targets in the mucosal immune system. In sexual mucosal transmission, these target cells at portals of entry enable these viruses to initially establish founder populations of infected cells that will seed and establish infection throughout the secondary lymphoid organ system. The GALT compartment probably has a particularly important role in viral production in early infection for two reasons: (1) its size (GALT comprises most (~60%) of the secondary lymphoid organ system<sup>1</sup>); and (2) the large number of recently activated memory CD4<sup>+</sup> T cells that, with current markers, appear to be resting but in fact are in a state in which they can be productively infected. We have shown that these resting cells contribute substantially to peak virus production, which incites immune activation to supply activated CD4<sup>+</sup> T cells to sustain infection. Furthermore, because lamina propria cells are at a 'tipping point' to undergo apoptosis as a regulatory mechanism to maintain the balance between host defence and the immunopathological consequences of prolonged activation of effector cells, exposure to large quantities of gp120 on virions triggers the apoptotic pathway on a correspondingly large scale, resulting in the preferential and nearly complete loss of CD4<sup>+</sup> T cells in the effector arm of GALT. The extent and rapidity of depletion of CD4<sup>+</sup> T cells in this major lymphatic tissue compartment in both SIV and HIV infections is in our view one major reason to focus on developing countermeasures to attack these viruses at or shortly after transmission, particularly in cervico-vaginal tissues, the globally predominant portal of entry for HIV-1 (ref. 23). □

## Methods

### SIV-infected animals

Adult female rhesus macaques (*Macaca mulatta*), housed at the California National Primate Research Center in accordance with the regulations of the American Association of Accreditation of Laboratory Animal Care standards, were inoculated intra-vaginally with  $2 \times 10^5$  50% tissue culture infectious dose (TCID<sub>50</sub>) of SIVmac251 or SIVmac239 (ref. 24). Colon tissue used in the studies described here was obtained from 14 out of 30 animals necropsied from 2 h to 28 days after inoculation.

### In situ hybridization and quantification of SIV RNA<sup>+</sup> cells

SIV RNA was detected in cells in formalin-fixed and paraffin-embedded tissues as previously described<sup>12,13</sup>. Briefly, 6–8-μm sections were cut and adhered to silanized slides. After de-paraffinization in xylene, rehydration in PBS and permeabilization by treating the sections with HCl, digitonin and proteinase K, the sections were acetylated and hybridized to <sup>35</sup>S-labelled SIV-specific riboprobes. After washing and digestion with RNases, sections were coated with nuclear track emulsion, exposed, developed and counterstained with Giemsa. SIV RNA-positive cells were automatically enumerated in sections of defined areas using MetaMorph software.

### Immunohistochemical staining and in situ hybridization

Combined immunohistochemical staining and *in situ* hybridization were performed as described previously<sup>12,13</sup>. Briefly, sections were microwaved for antigen retrieval, hybridized, washed and digested with RNases, incubated with antibody markers for cell type—naïve CD45RA<sup>+</sup> and CD45RO<sup>+</sup> memory CD4<sup>+</sup> T cells (anti-OPD4, ref. 25)—activation (CD69 and CD25) and proliferation (Ki67), and then stained with diaminobenzidine (DAB) with the the Dako Envision and Peroxidase kit. After washing, the sections were coated with nuclear track emulsion, exposed, developed and counterstained with haematoxylin.

### Immunofluorescence and CD4<sup>+</sup> and Ki67<sup>+</sup> CD4<sup>+</sup> cell counts

Sections were stained with primary antibodies to CD4, CD8, Ki67 and activated caspase-3, and secondary antibodies were labelled with Alexa Fluor 488 (green) and Alexa Fluor 555 (red). Cell nuclei were counterstained blue with TOTO-3. For quantification, sequential images at wavelengths for each fluorophore, collected using a Bio-Rad MRC 1000 Confocal Microscope at  $\times 60$ , were automatically processed with an action program in Adobe Photoshop 7.0. CD4<sup>+</sup> T cells and CD4<sup>+</sup> Ki67<sup>+</sup> cells were manually counted using a grid in Photoshop 7.0 in 30 serially captured images.

### Quantification of doubly labelled cells

The proportions of SIV RNA-positive Ki67<sup>+</sup> or Ki67<sup>-</sup> cells (Fig. 3b) were determined by enumerating >100 single- and double-labelled cells in sections after *in situ* hybridization and immunohistochemical staining to detect SIV RNA and Ki67. The number of cells per mm<sup>2</sup> (Fig. 3f) was determined from these proportions and the number of CD4<sup>+</sup> T cells per mm<sup>2</sup> determined as described above.



## In situ cell death detection by TUNEL assay

Endogenous peroxidase in tissue sections was blocked with 3% H<sub>2</sub>O<sub>2</sub> in methanol. Sections were then incubated at 98 °C in 100 mM citrate buffer, pH 6.0, cooled, immersed in 3% normal sheep serum in TNB blocking buffer (0.1 M TRIS-HCl, pH 7.5, 0.15 M NaCl, 0.5% blocking reagent (Dupont)) and incubated at 37 °C in a TUNEL reaction mixture (Roche). After rinsing, TUNEL-positive cells were stained with converter-POD reagent and DAB substrate.

Received 6 December 2004; accepted 7 March 2005; doi:10.1038/nature03513.  
Published online 27 March 2005.

1. Mowat, A. M. & Viney, J. L. The anatomical basis of intestinal immunity. *Immunol. Rev.* **156**, 145–166 (1997).
2. Veazey, R. S. *et al.* Gastrointestinal tract as a major site of CD4<sup>+</sup> T cell depletion and viral replication in SIV infection. *Science* **280**, 427–431 (1998).
3. Kewenig, S. *et al.* Rapid CD4<sup>+</sup> T cell depletion and enteropathy in simian immunodeficiency virus infected rhesus macaques. *Gastroenterology* **116**, 1115–1123 (1999).
4. Schneider, T. *et al.* Loss of CD4 T lymphocytes in patients infected with human immunodeficiency virus type 1 is more pronounced in the duodenal mucosa than in the peripheral blood. *Gut* **37**, 524–529 (1995).
5. Mehndru, S. *et al.* Primary HIV-1 infection is associated with preferential depletion of CD4<sup>+</sup> T lymphocytes from effector sites in the gastrointestinal tract. *J. Exp. Med.* **200**, 761–770 (2004).
6. Brechley, J. M. *et al.* CD4<sup>+</sup> T cell depletion during all stages of HIV disease occurs predominantly in the gastrointestinal tract. *J. Exp. Med.* **200**, 749–759 (2004).
7. Guadalupe, M. *et al.* Severe CD4<sup>+</sup> T-cell depletion in gut lymphoid tissue during primary human immunodeficiency virus type 1 infection and substantial delay in restoration following highly active antiretroviral therapy. *J. Virol.* **77**, 11708–11717 (2003).
8. Clayton, F., Snow, G., Reka, S. & Kotler, D. P. Selective depletion of rectal lamina propria rather than lymphoid aggregate CD4 lymphocytes in HIV infection. *Clin. Exp. Immunol.* **107**, 288–292 (1997).
9. Lim, S. G. *et al.* Loss of mucosal CD4 lymphocytes is an early feature of HIV infection. *Clin. Exp. Immunol.* **92**, 448–454 (1993).
10. Veazey, R. & Lackner, A. The mucosal immune system and HIV-1 infection. *AIDS Rev.* **5**, 245–252 (2003).
11. Veazey, R. *et al.* Identifying the target cell in primary simian immunodeficiency virus infection: Highly activated memory CD4<sup>+</sup> T cells are rapidly eliminated in early SIV infection *in vivo*. *J. Virol.* **74**, 57–64 (2000).
12. Zhang, Z.-Q. *et al.* Sexual transmission and propagation of simian and human immunodeficiency viruses in two distinguishable populations of CD4<sup>+</sup> T cells. *Science* **286**, 1353–1357 (1999).
13. Zhang, Z.-Q. *et al.* Roles of substrate availability and infection of resting and activated CD4<sup>+</sup> T cells in transmission and acute simian immunodeficiency virus infection. *Proc. Natl Acad. Sci. USA* **101**, 5640–5645 (2004).
14. Boirivant, M. *et al.* HIV-1 gp120 accelerates Fas-mediated activation-induced human lamina propria T cell apoptosis. *J. Clin. Immunol.* **18**, 39–47 (1998).
15. Boirivant, M. *et al.* Stimulated human lamina propria T cells manifest enhanced Fas-mediated apoptosis. *J. Clin. Invest.* **98**, 2616–2622 (1996).
16. De Maria, R. *et al.* Functional expression of Fas and Fas ligand on human gut lamina propria T lymphocytes. A potential role for the acidic sphingomyelinase pathway in normal immunoregulation. *J. Clin. Invest.* **97**, 316–322 (1996).
17. Testi, R., Phillips, J. H. & Lanier, L. L. Constitutive expression of a phosphorylated activation antigen (Leu 23) by CD3 bright human thymocytes. *J. Immunol.* **141**, 2557–2563 (1988).
18. Testi, R., Phillips, J. H. & Lanier, L. L. Leu 23 induction as an early marker of functional CD3/T cell antigen receptor triggering. Requirement for receptor cross-linking, prolonged elevation of intracellular [Ca<sup>++</sup>] and stimulation of protein kinase C. *J. Immunol.* **142**, 1854–1860 (1989).
19. Phillips, A. N. Reduction of HIV concentration during acute infection: independence from a specific immune response. *Science* **271**, 497–499 (1996).
20. Reilly, C. S. *et al.* The clustering of SIV infected cells in lymphatic tissue. *J. Am. Stat. Assoc.* **97**, 943–954 (2002).
21. Mothe, B. R. *et al.* Dominance of CD8 responses specific for epitopes bound by a single major histocompatibility complex class I molecule during the acute phase of viral infection. *J. Virol.* **76**, 875–884 (2002).
22. Reynolds, M. R. *et al.* The CD8<sup>+</sup> lymphocyte response to major immunodominant epitopes after vaginal exposure to SIV: too late and too little. *J. Virol.* (in the press).
23. Pope, M. & Haase, A. T. Transmission, acute HIV-1 infection and the quest for effective vaccines, microbicides and other strategies to prevent infection. *Nature Med.* **9**, 847–852 (2003).
24. Miller, C. J. *et al.* Intravaginal inoculation of rhesus macaques with cell-free simian immunodeficiency virus results in persistent or transient viremia. *J. Virol.* **68**, 6391–6400 (1994).
25. Poppema, S., Lai, R. & Visser, L. Monoclonal antibody OPD4 is reactive with CD45RO, but differs from UCHL1 by the absence of monocytic reactivity. *Am. J. Pathol.* **139**, 725–729 (1991).
26. Miller, C. J. *et al.* Propagation and dissemination of infection after vaginal transmission of SIV. *J. Virol.* (submitted).

**Acknowledgements** We thank R. Veazey, L. Picker, J. Lifson, D. Douek and M. Roederer for discussions; L. Compton, D. Lu, B. Vang, K. Bost and R. Dizon of the Immunology Core Laboratory and Primate Services Unit at the CNPRC for technical assistance; and T. Leonard and C. O'Neill for help in preparing the figures and manuscript. This work was supported by grants from the National Institute of Allergy and Infectious Diseases and from the National Center for Research Resources.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to A.T.H. (haase001@umn.edu).

## Clathrin is required for the function of the mitotic spindle

Stephen J. Royle<sup>1</sup>, Nicholas A. Bright<sup>2</sup> & Leon Lagnado<sup>1</sup>

<sup>1</sup>MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK  
<sup>2</sup>Cambridge Institute for Medical Research and Department of Clinical Biochemistry, University of Cambridge, Addenbrooke's Hospital, Hills Road, Cambridge CB2 2XY, UK

Clathrin has an established function in the generation of vesicles that transfer membrane and proteins around the cell<sup>1–4</sup>. The formation of clathrin-coated vesicles occurs continuously in non-dividing cells<sup>5</sup>, but is shut down during mitosis<sup>6</sup>, when clathrin concentrates at the spindle apparatus<sup>7,8</sup>. Here, we show that clathrin stabilizes fibres of the mitotic spindle to aid congression of chromosomes. Clathrin bound to the spindle directly by the amino-terminal domain of clathrin heavy chain. Depletion of clathrin heavy chain using RNA interference prolonged mitosis; kinetochore fibres were destabilized, leading to defective congression of chromosomes to the metaphase plate and persistent activation of the spindle checkpoint. Normal mitosis was rescued by clathrin triskelia but not the N-terminal domain of clathrin heavy chain, indicating that stabilization of kinetochore fibres was dependent on the unique structure of clathrin. The importance of clathrin for normal mitosis may be relevant to understanding human cancers that involve gene fusions of clathrin heavy chain.

The subcellular distribution of clathrin depended on the phase of the cell cycle<sup>7–9</sup> (Supplementary Fig. 1). During interphase, green fluorescent protein (GFP)-tagged clathrin light chain a (GFP-LCa) in NRK cells was associated with the Golgi apparatus and numerous puncta representing clathrin-coated pits and vesicles<sup>5</sup> (Fig. 1a). But during metaphase, clathrin localized to kinetochore fibres of the mitotic spindle<sup>10</sup> and possibly interpolar microtubules, but not astral microtubules (Fig. 1a, b). Localization of clathrin to kinetochore fibres was confirmed by chilling cells for 10 min at 4 °C to selectively disassemble microtubules not associated with kinetochores<sup>11</sup>; after this treatment, clathrin in metaphase cells remained bound to the kinetochore fibres, indicating that these microtubules were a potential site of clathrin function (Fig. 1b). Similar changes in the distribution of clathrin were observed using other variants of the light chain tagged with GFP or by immunocytochemistry using a monoclonal antibody specific for clathrin heavy chain (CHC; Supplementary Figs 2 and 3).

Two observations indicated that clathrin bound to the mitotic spindle rather than membrane localized to this region. First, none of the major adaptor proteins that allow clathrin to coat membranes (AP-1, AP-2 and AP-3)<sup>2,3</sup> was found at the spindle apparatus (Supplementary Fig. 4a–c). To test whether clathrin at the spindle was associated with membranes at all, we indiscriminately labelled intracellular compartments by incubating cells with the styryl dye FM4-64 (15 μM) for >24 h (Supplementary Fig. S4d). In cells at metaphase, none of these membranes was found at the spindle (Fig. 1c).

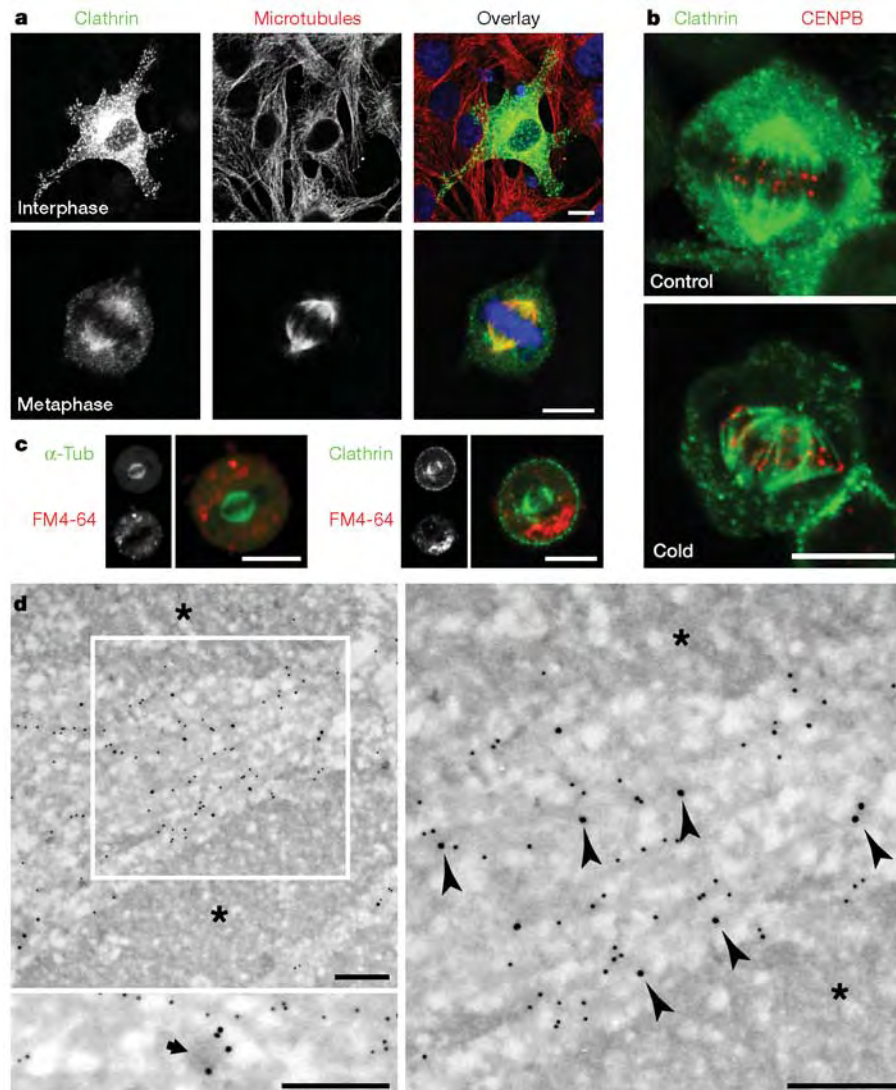
The localization of clathrin to the mitotic spindle was examined at higher resolution using immunoelectron microscopy. CHC and α-tubulin were immunolabelled with 15-nm and 10-nm colloidal-gold-conjugated reagents, respectively (Fig. 1d). CHC in mitotic NRK cells was associated with electron-dense tracks that were directed towards the chromosomes, and these tracks were confirmed as microtubules by labelling for α-tubulin (Fig. 1d, top left and right panels). Clathrin-coated vesicles were scarce in mitotic cells but when visualized (Fig. 1d, bottom left panel) they could be distinguished from microtubule-associated clathrin (Fig. 1d, right

panel). Furthermore, clathrin immunoreactivity persisted at the spindle after soluble proteins had been removed from cells by detergent extraction (Supplementary Fig. 3). Together, these results indicated that clathrin at the mitotic spindle was not coating membranes but was bound to microtubules or microtubule-associated protein(s). Direct binding of CHC to the spindle apparatus has also been demonstrated by mass spectrometry<sup>12</sup>.

To identify the region of clathrin that determined its association with the mitotic spindle, recruitment was quantified using a simple assay that compared the intensity of fluorescent proteins in the region of the spindle relative to the cytoplasm (see Methods). GFP-LCa was recruited to the spindle when CHC was abundant in cells, but not when CHC was depleted by RNA interference (RNAi), indicating that the determinant for spindle binding was contained in the heavy chain (Fig. 2a, b). To localize the region further, GFP-tagged fragments of CHC were expressed and their recruitment to the spindle quantified (Fig. 2c–e). CHC fragments

containing the N-terminal domain (GFP-CHC(1–1074) and GFP-CHC(1–479)) and the N-terminal domain itself (GFP-CHC(1–330)) were recruited, but a long construct lacking this region (GFP-CHC(331–1074)) was not (Fig. 2d, e). These results indicate that clathrin triskelia bound to the mitotic spindle via the N-terminal domain of the heavy chain.

Having characterized how clathrin becomes associated with the mitotic spindle, we went on to investigate whether it had any role in mitosis by using RNAi to deplete rat or human cells of CHC (see Supplementary Information). In NRK cells 72 h after transfection, the level of CHC at interphase was ~10% of controls and clathrin-mediated endocytosis was reduced by more than 90% (Supplementary Fig. 5). At metaphase, the level of CHC at the spindle was  $11 \pm 2\%$  of control levels 72 h after transfection with short interfering (si)RNA. RNAi-mediated depletion of CHC also reduced proliferation of cells twofold, even though the proportion of dead or dying cells was only ~0.3%, as judged by nuclear morphology<sup>13,14</sup>.



**Figure 1** Clathrin was targeted to the mitotic spindle of NRK cells. **a**, Confocal micrographs showing the subcellular distribution of clathrin at interphase and metaphase. GFP-LCa (left, green),  $\alpha$ -tubulin (centre, red) and nucleic acid (blue) staining is shown. **b**, Cells expressing GFP-LCa fixed before (top) or after (bottom) cold treatment to depolymerize non-kinetochore microtubules. CENPB, centromere protein B. **c**, **d**, The association of clathrin with microtubules is not via coated membranes. **c**, Example images of live cells expressing either GFP- $\alpha$ -tubulin (left) or GFP-LCa (right) imaged after

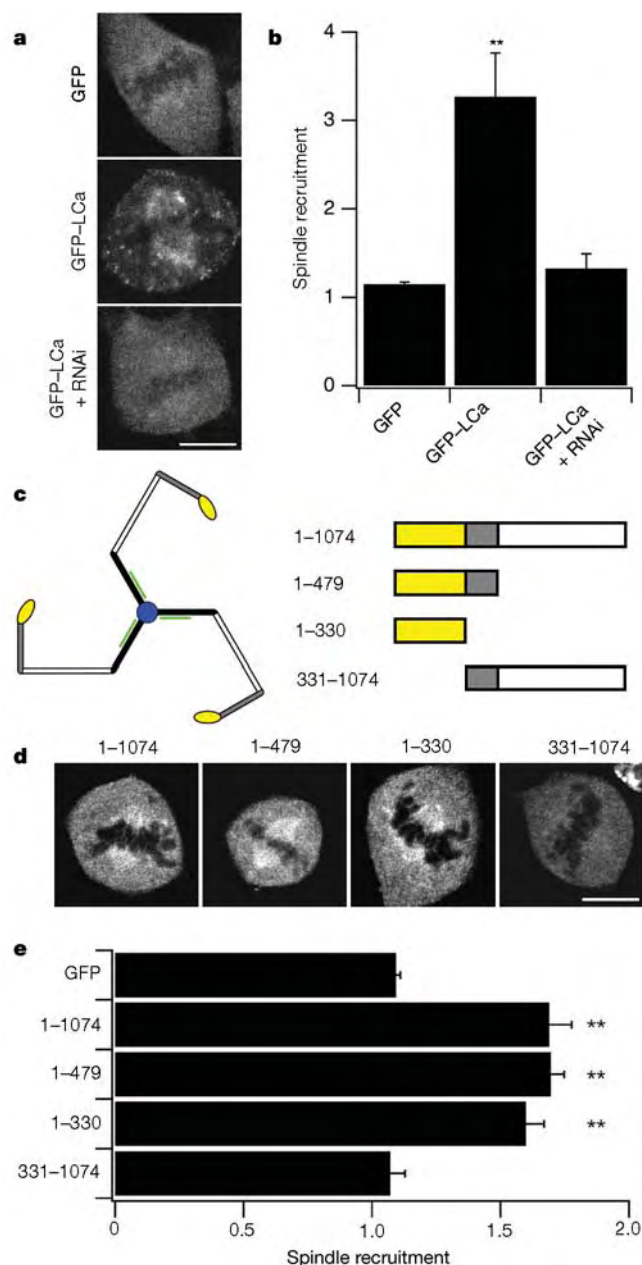
24–28-h incubation with FM4-64 (red). **d**, Association of clathrin with microtubules visualized by immunogold electron microscopy. CHC (15 nm gold) and  $\alpha$ -tubulin (10 nm gold) in mitotic NRK cells. Chromosomes are denoted by asterisks. A morphologically distinct clathrin-coated vesicle (bottom left panel) is indicated by an arrow. Arrowheads denote CHC labelling associated with microtubules. Scale bars: 10  $\mu$ m (**a–c**); 250 nm (**d**).

Reduced proliferation was associated with prolonged mitosis, because the proportion of cells in mitosis (the mitotic index) was increased fourfold 72 h after CHC RNAi transfection (rat: control  $3.5 \pm 0.2\%$ , knockdown  $14.2 \pm 1.1\%$  (Fig. 3a); human: control  $1.9 \pm 0.21\%$ , knockdown  $8.4 \pm 0.6\%$ ;  $P < 0.001$ ; data not shown).

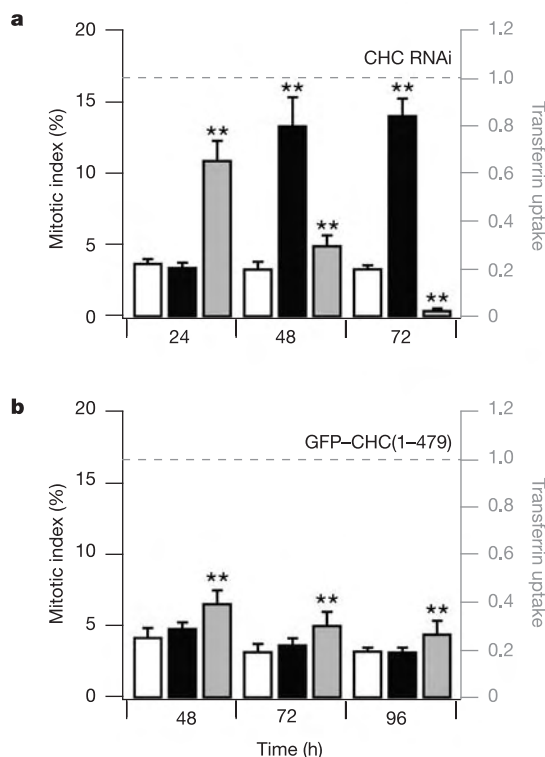
Does prolonged mitosis in cells depleted of CHC represent an alternative function of clathrin or an indirect effect of inhibiting clathrin-mediated endocytosis? To investigate this question a dominant-negative inhibitor of clathrin-mediated endocytosis was used as an alternative method of inhibiting the established function of

clathrin. When GFP-CHC(1–479) was overexpressed in NRK cells, uptake of fluorescent transferrin was inhibited by 60–70% during a 48–96 h period after transfection, but the mitotic index was not affected (Fig. 3b). Overexpression of GFP-CHC(1–479) inhibited clathrin-mediated endocytosis without significantly affecting the binding of clathrin to the spindle: spindles in GFP-CHC(1–479)-expressing cells had  $94 \pm 6\%$  of the clathrin immunoreactivity observed in cells expressing GFP alone. The prolongation of mitosis in cells depleted of clathrin therefore reflected a direct action of clathrin at the mitotic spindle, distinct from its role in membrane transport. This conclusion was reinforced by two observations. First, clathrin-mediated endocytosis was shut down during mitosis (transferrin uptake in mitotic cells was  $10 \pm 6\%$  of that at interphase)<sup>6</sup>. Second, knockdown of CHC caused a series of mitotic defects (described below), none of which was observed in cells in which clathrin-mediated endocytosis was inhibited by expression of GFP-CHC(1–479).

A number of observations indicated that clathrin regulated the congression of chromosomes. First, when we examined the proportion of cells at each stage of mitosis, more were in prometaphase after clathrin knockdown compared with controls ( $49 \pm 6\%$  versus  $31 \pm 6\%$ ,  $P < 0.05$ ; Supplementary Fig. 5f). Second, the metaphase plate was thicker in clathrin-depleted cells ( $11.2 \pm 0.9 \mu\text{m}$  compared with  $7.0 \pm 0.5 \mu\text{m}$  in controls; Fig. 4a) and centromeres did not organize on the mitotic spindle in an orderly manner (Fig. 4b). Third, there was an increased incidence of misaligned chromosomes



**Figure 2** Clathrin was targeted to the mitotic spindle via the N-terminal domain of the heavy chain. **a**, Example images of GFP and GFP-LCa in cells at metaphase. **b**, Histogram of spindle recruitment of GFP or GFP-LCa. A value of one represents no specific recruitment (see Methods). **c**, Schematic diagram of a clathrin triskelion and the CHC fragments used in **d**, **e**. The CHC N-terminal domain forming the 'foot' is yellow (residues 1–330). **d**, Example images of GFP-tagged CHC fragments in cells at metaphase. **e**, Histogram of spindle recruitment of GFP or GFP-tagged CHC fragments. Results are mean  $\pm$  s.e.m.; double asterisk,  $P < 0.01$ . Scale bars:  $10 \mu\text{m}$ .



**Figure 3** Inhibition of clathrin-mediated endocytosis did not disrupt mitosis. **a**, Effect of clathrin depletion on mitotic index (black bars) and transferrin uptake (grey bars) 24, 48 and 72 h after transfection with CHC siRNA. Open bars show mitotic index in control siRNA-transfected cells. Transferrin uptake was normalized to control values (dotted line). Within 48 h clathrin-mediated endocytosis was reduced by 70% and the mitotic index increased fourfold. **b**, Effects of inhibiting clathrin-mediated endocytosis by overexpression of GFP-CHC(1–479), measured 48, 72 and 96 h after transfection. Transferrin uptake was inhibited by 60–70%, without any change in the mitotic index. Control (open bars) was GFP alone. Results are mean  $\pm$  s.e.m.; double asterisk,  $P < 0.01$ .

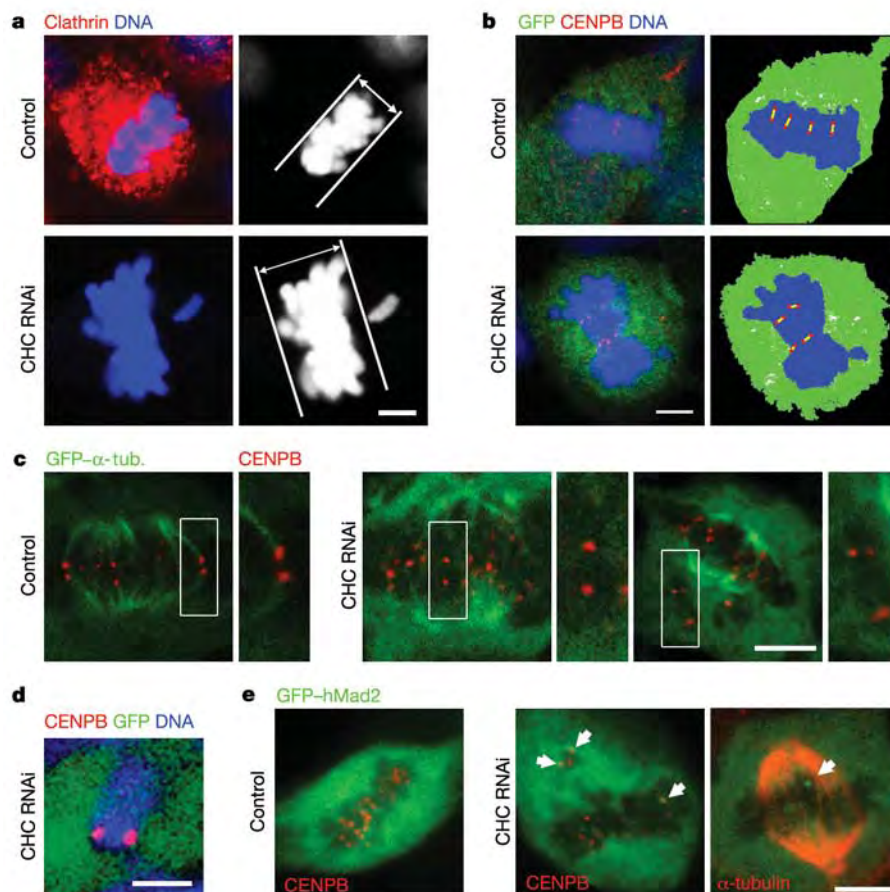


in metaphase-like cells after clathrin knockdown (Fig. 4a; see also the right CHC RNAi panel of c and e for other examples). Misaligned chromosomes were normally observed in  $4 \pm 1\%$  of rat cells at metaphase and  $10 \pm 7\%$  of human cells. However, after depletion of clathrin, misaligned chromosomes were observed in  $22 \pm 5\%$  of rat cells and  $70 \pm 7\%$  of human cells that appeared to be in metaphase. In these human cells, there was an average of  $3.4 \pm 0.8$  misaligned chromosomes per cell. Misaligned chromosomes were usually found at spindle poles; they always consisted of pairs of sister chromatids (Fig. 4d) but they did not have spindle attachments (Fig. 4c, right CHC RNAi panel) and the arms were very rarely in the V-shaped configuration typical of congressing chromosomes<sup>15</sup>. Misaligned chromosomes therefore arose owing to a failure in congression during prometaphase rather than premature separation of sister chromatids.

Kinetochores fibres exert tension on sister chromatids that can be assessed by measuring the interkinetochore distance<sup>16</sup>. At early prometaphase, before chromosomes attach to the spindle, this distance was  $0.8 \pm 0.02 \mu\text{m}$  in controls and  $0.8 \pm 0.02 \mu\text{m}$  in cells depleted of clathrin. The upper 95% confidence interval of the control distribution was  $1.2 \mu\text{m}$ , so interkinetochore distances greater than this threshold value could be taken as evidence that the sister chromatids were under tension. Misaligned chromosomes in clathrin-depleted cells at metaphase were not under tension, because the interkinetochore distance averaged  $0.9 \pm 0.05 \mu\text{m}$

( $P = 0.151$ ). At metaphase, the interkinetochore distance of equatorial chromosomes averaged  $1.6 \pm 0.04 \mu\text{m}$  in control cells, but was significantly less in cells depleted of clathrin ( $1.3 \pm 0.03 \mu\text{m}$ ;  $P < 0.01$ ). Notably, only 2% of kinetochore pairs were at a distance less than  $1.2 \mu\text{m}$  at the metaphase plate of control cells, but this reached 20% in cells depleted of clathrin. These results indicate that clathrin knockdown led to a reduction in tension exerted on sister chromatids of some chromosomes. Clathrin knockdown also reduced the stability of kinetochore–spindle contacts. In control cells at metaphase, all kinetochores had an attached fibre, as judged by the selective depolymerization of microtubules not attached to kinetochores (Fig. 4c). In contrast, clathrin-depleted cells often contained ‘orphan’ centromeres that did not have a fibre attached (Fig. 4c, left CHC RNAi panel), suggesting that congression had occurred but that the fibre had been lost subsequently.

The transition from metaphase to anaphase is controlled by the spindle checkpoint, which monitors correct attachment of chromosomes to kinetochore fibres<sup>17</sup>. Persistent activation of the checkpoint prolongs mitosis<sup>17</sup>. One component of the checkpoint is Mad2, which we visualized by coupling expression of GFP–hMad2 (ref. 18) to either control or CHC short hairpin RNA (shRNA) in human cells using pBrain vectors (see Methods). In cells co-expressing control shRNA, GFP–hMad2 correctly localized to kinetochores in early prometaphase (not shown) and then became diffusely distributed at metaphase<sup>18</sup> (Fig. 4e, left panel).



**Figure 4** Depletion of clathrin results in destabilized kinetochore fibres, defective congression of chromosomes and prolonged activation of the spindle checkpoint. **a**, Clathrin depletion increased the frequency of misaligned chromosomes and caused thicker metaphase plates (parallel lines). **b**, Clathrin-depleted metaphase-like plates were disorganized. Cells were marked by GFP (green) and stained for nucleic acids (blue) and CENPB (red) to visualize centromeres. Schematic drawings illustrate centromere

arrangement (right). **c**, Cells expressing GFP– $\alpha$ -tubulin (green) after depolymerization of non-kinetochore fibres. The right panels show a higher magnification of boxed centromere pairs. **d**, Misaligned chromosomes in CHC RNAi cells were pairs of sister chromatids. **e**, Representative images of one control (left) and two CHC RNAi (middle and right panels) cells expressing very low levels of GFP–hMad2. Mad2-positive kinetochores are indicated by arrows. Scale bars:  $5 \mu\text{m}$ .

But in cells co-expressing CHC shRNA, Mad2 signalling persisted during metaphase: GFP-hMad2 was found on the kinetochores of misaligned chromosomes as well as chromosomes at the metaphase plate (Fig. 4e). Staining for  $\alpha$ -tubulin revealed that Mad2-positive kinetochores had no obvious microtubular connections (Fig. 4e,

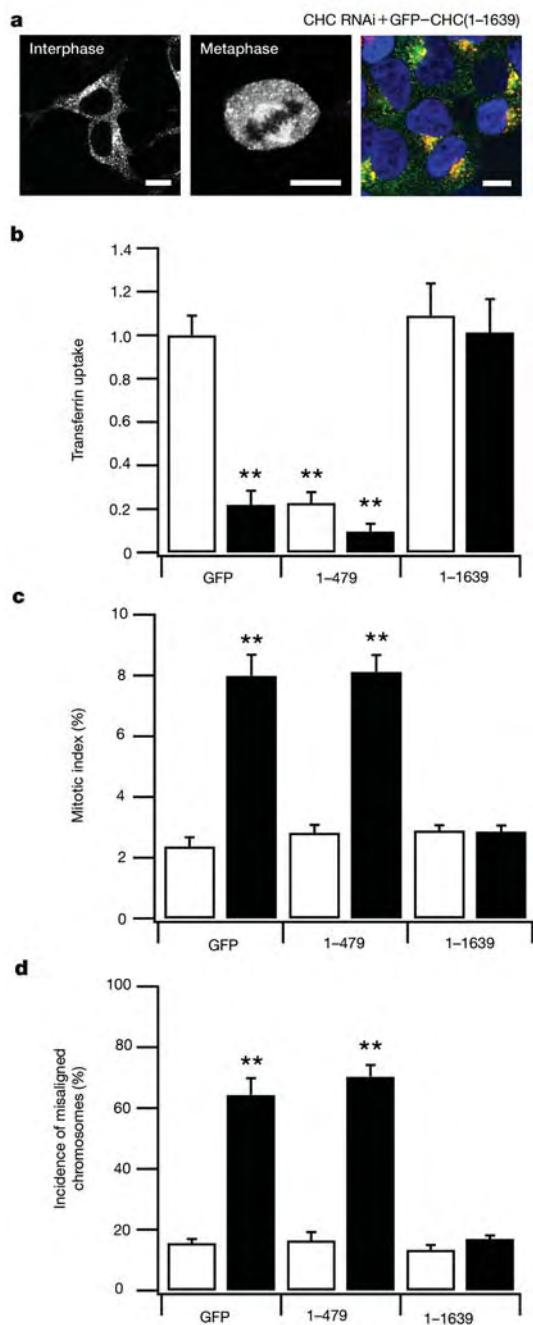
right panel). Similar observations were made using anti-Mad2 (Supplementary Fig. 7b, c). These results indicate that the primary cause of prolonged mitosis in cells depleted of clathrin was the continued activation of the spindle checkpoint, which resulted from destabilization of kinetochore microtubules.

How might clathrin influence the stability of kinetochore fibres? Free clathrin occurs as triskelia<sup>4,19</sup>, and the N-terminal domain at the foot of each leg, which is required for binding to the spindle (Fig. 2), also interacts with a large number of proteins<sup>20,21</sup>. One possibility is that clathrin triskelia stabilize kinetochore fibres by acting as a brace connecting two or three microtubules within a fibre. Alternatively, the N-terminal domain might simply recruit another protein required for stabilization of spindle fibres. To distinguish between these two possibilities, we tested whether clathrin triskelia or the N-terminal domain alone could rescue the mitotic defects in cells depleted of endogenous clathrin. For these experiments we expressed either GFP (control), GFP-CHC(1-479) or full-length GFP-CHC(1-1639) in control or CHC RNAi cells, while making the CHC components resistant to RNAi by silent mutations in the shRNA-binding region (see Methods). For full-length CHC we used the major human splice variant that encodes residues 1-1639. Two observations indicated that GFP-CHC(1-1639) formed triskelia with normal function in clathrin-mediated endocytosis. First, this construct labelled punctate structures in interphase cells and decorated the mitotic spindle at metaphase (Fig. 5a), similar to GFP-LCa<sup>5</sup> (Fig. 1). Second, GFP-CHC(1-1639) supported the uptake of transferrin in a manner indistinguishable from wild-type CHC (Fig. 5a, b). GFP-CHC(1-1639) also corrected mitotic defects in cells depleted of endogenous clathrin, preventing the prolongation of mitosis (Fig. 5c) and returning the incidence of misaligned chromosomes to normal levels (Fig. 5d). In contrast, the N-terminal domain alone had no significant effect on mitigating these consequences of clathrin knockdown (Fig. 5c, d). We conclude that stabilization of kinetochore fibres requires the trimeric structure of clathrin rather than the interaction function of the N-terminal domain.

Kinetochore fibres might be strengthened if clathrin triskelia form a relatively rigid connection between microtubules. In agreement with this idea, electron micrographs show that microtubules within the spindle fibres are connected by curved 'bridges'<sup>10,22</sup>, the molecular identities of which are currently unknown. These bridges result in 50–100-nm spacing between microtubules in spindle fibres<sup>10</sup>. By comparison, the distance between N-terminal domains of free triskelia is 45–70 nm<sup>19,23</sup>.

We found an increased frequency of misaligned chromosomes in cells depleted of clathrin (Figs 4 and 5). The mis-segregation of chromosomes during mitosis is a potential source of aneuploidy, a form of genetic instability that may lead to cancer or birth defects<sup>24</sup>. Given the evidence that clathrin has an alternative function in mitosis, it may be worth re-assessing the involvement of gene fusions of CHC with anaplastic lymphoma kinase (ALK)<sup>25</sup> in inflammatory myofibroblastic tumours and anaplastic large-cell lymphoma, as well as the fusion of CHC with the transcription factor gene *TFE3* in renal adenocarcinomas<sup>26</sup>. These gene fusions occur at the carboxy-terminal end of CHC, and are therefore expected to disrupt trimerization while allowing binding to the mitotic spindle. On the basis of our results, we suggest that these fusion proteins might impair the function of clathrin during mitosis, or in the case of CHC-ALK, target a catalytically active fragment of ALK to the mitotic spindle.

The role of clathrin in the transport of membrane and proteins has been studied intensively<sup>1-3</sup>. Here, we have provided evidence for a second important function of clathrin, which occurs during mitosis; that is, the stabilization of kinetochore fibres of the spindle apparatus. Future studies will seek to identify protein partners for clathrin at the spindle and understand how clathrin switches between its two functions. □



**Figure 5** Full-length CHC, but not CHC N-terminal domain, is sufficient to rescue the mitotic defects found in cells depleted of endogenous CHC. **a**, Representative images of GFP-tagged knockdown-resistant CHC (GFP-CHC(1-1639)) expressed in HEK293 cells that were depleted of endogenous CHC. The right panel shows the normal uptake of transferrin (red), GFP-CHC(1-1639) (green) and DNA (blue) in these cells. **b–d**, Quantification of transferrin uptake at interphase (**b**), mitotic index (**c**) and the frequency of metaphase-like cells with misaligned chromosomes (**d**) in cells expressing GFP, GFP-CHC(1-479) or GFP-CHC(1-1639) 72 h after transfection. Open bars, control cells; filled bars, cells depleted of endogenous CHC. Results are mean  $\pm$  s.e.m.; double asterisk,  $P < 0.01$ .

## Methods

### Molecular biology

GFP-LCa was generated by polymerase chain reaction to introduce *Bgl*II and *Eco*RI sites, and subcloned into pEGFP-C1 (Clontech). GFP-CHC(1–479), GFP-CHC(1–330) and GFP-CHC(331–1074) were amplified and subcloned into *Bgl*II and *Hind*III sites of pEGFP-C1, and GFP-CHC(1–1074) was made by subcloning a *Bgl*II–*Sca*I fragment from GFP-CHC(1–479) into GFP-CHC(331–1074). GFP-hMad2 was reconstructed to enable us to make pBrain versions (see below). GFP-hMad2 in pCS2 was amplified to introduce *Bgl*II and *Hind*III sites, and the resulting fragment was cloned into pEGFP-C1. Human CHC and clathrin light chain LCa complementary DNAs (I.M.A.G.E. 6187185 and 3944942) were purchased from MRC Geneservice. GFP-tagged  $\alpha$ -tubulin was from Clontech (pEGFP-Tub). GFP-hMad2 in pCS2 (ref. 18) was a gift from G. Fang. A series of vectors (pBrain) were made that allowed the simultaneous expression of shRNA under an H1 RNA promoter and fluorescent proteins under a CMV promoter. For rescue experiments, GFP-CHC(1–479) was rendered resistant to knockdown ('knockdown-proof'; KDP) by mutation using the megaprimer method to give GFP-CHC(1–479)KDP. Knockdown-proof GFP-CHC(1–1639) was made by subcloning an *Age*I–*Mfe*I fragment into pEGFP-C1 at *Xma*I–*Mfe*I sites and then repairing the N terminus by substituting a *Bgl*II–*Asp*718 fragment from GFP-CHC(1–479)KDP. All constructs used in this study were verified by automated DNA sequencing (Lark or MRC Geneservice).

### Immunocytochemistry

Immunostaining was performed as described previously<sup>27</sup>. The following monoclonal antibodies were used: anti-clathrin heavy chain and anti- $\alpha$ -adaptin (X22 and AP6, Affinity BioReagents); anti- $\alpha$ -tubulin and anti- $\beta$ 1/2-adaptin (DM1A and 100/1, Sigma); anti-CENPB (a gift from W. C. Earnshaw); and anti- $\delta$ -adaptin (clone SA4; a gift from M. S. Robinson). Rabbit polyclonal anti-clathrin antiserum was as previously described<sup>28</sup>. Goat anti-mouse or anti-rabbit Cy3-conjugated secondary antibodies were from Jackson ImmunoResearch. Goat anti-mouse IgG conjugated to 10-nm colloidal gold was from BioCell. Protein A conjugated to 15-nm colloidal gold was from the Department of Cell Biology, University of Utrecht. TOPRO-3 (Molecular Probes) and Hoechst 33342 (Sigma) were used for staining DNA/RNA. Uptake of transferrin-Alexa546 (Molecular Probes) was done as previously described<sup>29</sup>. FM4-64 was from Calbiochem. For membrane-labelling experiments, transfected cells were cultured for >24 h in 15  $\mu$ M FM4-64 at 37 °C, and cells were washed for 5 min in imaging buffer (MEM without phenol red, 10% FBS, 100 U ml<sup>–1</sup> penicillin/streptomycin) before images were taken.

### Imaging

Confocal imaging was done using a BioRad Radiance 2000 and Nikon TE300 microscope with  $\times 60$  (1.4 NA) or  $\times 100$  (1.3 NA) oil immersion objectives. GFP, Cy3 or FM4-64, and TOPRO-3 were excited at 488, 543 and 633 nm, respectively. For quantitative immunostaining experiments, identical laser power and acquisition settings were used. Images (8-bit) were imported into IMAGEJ (NIH) or IPLab 3.9 (Scanalytics). To quantify the uptake of transferrin-Alexa546, the outline of the cell was drawn on the GFP channel of the image and this region of interest (ROI) was transferred to the red channel, the image was assigned a threshold and the number of transferrin puncta counted. Spindle recruitment was assayed by dividing the mean pixel density measured in a  $1 \times 1 \mu$ m ROI placed over the spindle by that measured in a region outside the spindle.

The mitotic index was determined by counting the number of cells in mitosis as a proportion of the total number of cells within a  $275 \times 190 \mu$ m area. For image quantification and counting experiments, between 5–80 cells were analysed and 100–3,914 cells were counted from experiments performed three to six times. Results are expressed as mean  $\pm$  s.e.m.; unpaired Student's *t*-test was used to compare control and test values, and analysis of variance (ANOVA) with Dunnett's post-hoc test was used to compare multiple groups to a control. For binomial results (mitotic index, misaligned chromosomes, multinucleate cells, and so on) data were tested for approximation to a normal distribution, *z*-values were calculated and *P*-values were retrieved using Microsoft Excel.

### Immunoelectron microscopy

Cells were prepared for ultra-structural analysis using immunogold electron microscopy as previously described<sup>30</sup>. Briefly, mitotic NRK cells were fixed with 4% paraformaldehyde/0.1% glutaraldehyde in 0.1 M sodium cacodylate, pH 7.2 at room temperature for 1 h, infused with 1.7 M sucrose/15% polyvinylpyrrolidone and prepared as previously described<sup>30</sup>. Ultra-thin frozen sections were collected from the knife edge with 50:50 2% methyl cellulose:2.3 M sucrose and immunolabelled, contrasted with methyl cellulose/uranyl acetate, dried and observed in a Philips CM100 TEM<sup>30</sup>. (See Supplementary Information for a complete description of the Methods.)

Received 7 January; accepted 1 March 2005; doi:10.1038/nature03502.

1. Kirchhausen, T. Clathrin. *Annu. Rev. Biochem.* **69**, 699–727 (2000).
2. Brodsky, F. M., Chen, C. Y., Knuehl, C., Towler, M. C. & Wakeham, D. E. Biological basket

- weaving: formation and function of clathrin-coated vesicles. *Annu. Rev. Cell Dev. Biol.* **17**, 517–568 (2001).
3. Robinson, M. S. Adaptable adaptors for coated vesicles. *Trends Cell Biol.* **14**, 167–174 (2004).
4. Fotin, A. *et al.* Molecular model for a complete clathrin lattice from electron cryomicroscopy. *Nature* **432**, 573–579 (2004).
5. Gaidarov, I., Santini, F., Warren, R. A. & Keen, J. H. Spatial control of coated-pit dynamics in living cells. *Nature Cell Biol.* **1**, 1–7 (1999).
6. Warren, G. Membrane partitioning during cell division. *Annu. Rev. Biochem.* **62**, 323–348 (1993).
7. Maro, B., Johnson, M. H., Pickering, S. J. & Louvard, D. Changes in the distribution of membranous organelles during mouse early development. *J. Embryol. Exp. Morphol.* **90**, 287–309 (1985).
8. Okamoto, C. T., McKinney, J. & Jeng, Y. Y. Clathrin in mitotic spindles. *Am. J. Physiol. Cell Physiol.* **279**, C369–C374 (2000).
9. Sutherland, H. G. *et al.* Large-scale identification of mammalian proteins localized to nuclear sub-compartments. *Hum. Mol. Genet.* **10**, 1995–2011 (2001).
10. Compton, D. A. Spindle assembly in animal cells. *Annu. Rev. Biochem.* **69**, 95–114 (2000).
11. Yao, X., Abrieu, A., Zheng, Y., Sullivan, K. F. & Cleveland, D. W. CENP-E forms a link between attachment of spindle microtubules to kinetochores and the mitotic checkpoint. *Nature Cell Biol.* **2**, 484–491 (2000).
12. Mack, G. J. & Compton, D. A. Analysis of mitotic microtubule-associated proteins using mass spectrometry identifies astrin, a spindle-associated protein. *Proc. Natl Acad. Sci. USA* **98**, 14434–14439 (2001).
13. Motley, A., Bright, N. A., Seaman, M. N. & Robinson, M. S. Clathrin-mediated endocytosis in AP-2-depleted cells. *J. Cell Biol.* **162**, 909–918 (2003).
14. Hinrichsen, L., Harborth, J., Andrees, L., Weber, K. & Ungewickell, E. J. Effect of clathrin heavy chain- and  $\alpha$ -adaptin-specific small inhibitory RNAs on endocytic accessory proteins and receptor trafficking in HeLa cells. *J. Biol. Chem.* **278**, 45160–45170 (2003).
15. Skibbens, R. V., Skeen, V. P. & Salmon, E. D. Directional instability of kinetochore motility during chromosome congression and segregation in mitotic newt lung cells: a push-pull mechanism. *J. Cell Biol.* **122**, 859–875 (1993).
16. Waters, J. C., Skibbens, R. V. & Salmon, E. D. Oscillating mitotic newt lung cell kinetochores are, on average, under tension and rarely push. *J. Cell Sci.* **109**, 2823–2831 (1996).
17. Cleveland, D. W., Mao, Y. & Sullivan, K. F. Centromeres and kinetochores: from epigenetics to mitotic checkpoint signaling. *Cell* **112**, 407–421 (2003).
18. Howell, B. J., Hoffman, D. B., Fang, G., Murray, A. W. & Salmon, E. D. Visualization of Mad2 dynamics at kinetochores, along spindle fibers, and at spindle poles in living cells. *J. Cell Biol.* **150**, 1233–1250 (2000).
19. Ungewickell, E. & Branton, D. Assembly units of clathrin coats. *Nature* **289**, 420–422 (1981).
20. ter Haar, E., Harrison, S. C. & Kirchhausen, T. Peptide-in-groove interactions link target proteins to the  $\beta$ -propeller of clathrin. *Proc. Natl Acad. Sci. USA* **97**, 1096–1100 (2000).
21. Miele, A. E., Watson, P. J., Evans, P. R., Traub, L. M. & Owen, D. J. Two distinct interaction motifs in amphiphysin bind two independent sites on the clathrin terminal domain  $\beta$ -propeller. *Nature Struct. Mol. Biol.* **11**, 242–248 (2004).
22. Hepler, P. K., McIntosh, J. R. & Cleland, S. Inter-microtubule bridges in mitotic spindle apparatus. *J. Cell Biol.* **45**, 438–444 (1970).
23. Kirchhausen, T., Harrison, S. C. & Heuser, J. Configuration of clathrin trimers: evidence from electron microscopy. *J. Ultrastruct. Mol. Struct. Res.* **94**, 199–208 (1986).
24. Jallepalli, P. V. & Lengauer, C. Chromosome segregation and cancer: cutting through the mystery. *Nature Rev. Cancer* **1**, 109–117 (2001).
25. Pulford, K., Morris, M. & Turturro, F. Anaplastic lymphoma kinase proteins in growth control and cancer. *J. Cell. Physiol.* **199**, 330–358 (2004).
26. Argani, P. *et al.* A novel CLTC-TFE3 gene fusion in pediatric renal adenocarcinoma with t(X;17)(p11.2;q23). *Oncogene* **22**, 5374–5378 (2003).
27. Bobanovic, L. K., Royle, S. J. & Murrell-Lagnado, R. D. P2X receptor trafficking in neurons is subunit specific. *J. Neurosci.* **22**, 4814–4824 (2002).
28. Simpson, F. *et al.* A novel adaptor-related protein complex. *J. Cell Biol.* **133**, 749–760 (1996).
29. Royle, S. J., Bobanovic, L. K. & Murrell-Lagnado, R. D. Identification of a non-canonical tyrosine-based endocytic motif in an ionotropic receptor. *J. Biol. Chem.* **277**, 35378–35385 (2002).
30. Bright, N. A., Reeves, B. J., Mullock, B. M. & Luzio, J. P. Dense core lysosomes can fuse with late endosomes and are re-formed from the resultant hybrid organelles. *J. Cell Sci.* **110**, 2027–2040 (1997).

**Supplementary Information** accompanies the paper on [www.nature.com/nature](http://www.nature.com/nature).

**Acknowledgements** We thank W. C. Earnshaw, G. Fang, G. Ihrke, A. P. Jackson and M. S. Robinson for their gifts of antibodies, plasmids and cells. We also thank J. W. Raff and M. S. Robinson for useful discussion. This work was supported by the MRC and the Human Frontiers Science Program (grant to L.L.). N.A.B. was funded by the MRC.

**Competing interests statement** The authors declare that they have no competing financial interests.

**Correspondence** and requests for materials should be addressed to S.J.R. (sjr51@mrc-lmb.cam.ac.uk).



## Contacts

**Publisher:** Ben Crowe

**Editor:** Paul Smaglik

**Marketing Manager:** David Bowen

### European Head Office, London

The Macmillan Building  
4 Crinan Street  
London N1 9XW, UK  
Tel +44 (0) 20 7843 4961  
Fax +44 (0) 20 7843 4996  
e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

### Naturejobs Sales Director:

Nevin Bayoumi (4978)

### European Sales Manager:

Andy Douglas (4975)

### UK/ RoW/ Ireland:

Nils Moeller (4953)  
Irene Vigila-Alton (4944)

### Scandinavia/ Spain/ Portugal:

Evelina Rubio Håkansson (4973)

**Natureevents:** Silke Opstrup (4994)

### France/ Switzerland:

Amelie Pequignot (4974)

### Advertising Production

**Manager:** Billie Franklin

To send materials use London  
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

### Naturejobs web development:

Tom Hancock

### Naturejobs online production:

Niamh Shields

### European Satellite Office

**Germany/ Austria/ Italy/**

**The Netherlands/ Belgium:**

Patrick Phelan

Tel +49 89 54 90 57 11

Fax +49 89 54 90 57 20

e-mail: [p.phelan@nature.com](mailto:p.phelan@nature.com)

Reya Silao

Tel +44 (0) 20 7843 4970

e-mail: [rsilao@nature.com](mailto:rsilao@nature.com)

### US Head Office, New York

345 Park Avenue South,  
10th Floor, New York, NY 10010-1707  
Tel +1 800 989 7718  
Fax +1 800 989 7103  
e-mail: [naturejobs@nature.com](mailto:naturejobs@nature.com)

### US Sales Manager:

Peter Bless

### Japan Head Office, Tokyo

MG Ichigaya Building (5F),  
19-1 Harakakamachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

### Asia-Pacific Sales Director:

Rinoko Asami

e-mail: [rasami@nature.com](mailto:rasami@nature.com)

# naturejobs

## A question of age

European universities and research agencies are increasingly looking beyond the continent for postdocs. This is, overall, a positive development. It makes European science more international and creates a pathway for researchers on the move. But some schemes, intended to promote mobility, unintentionally manage to restrict it.

For example, one listing for a non-German postdoc at a German university put an age limit of 29 years on the prospective fellow. On the face of it, this seems pretty innocuous, as Germans traditionally finish their PhDs at a relatively young age. But by placing an age restriction on candidates from other countries with different career paths and timelines, it may exclude the very people it is trying to attract.

The criterion almost automatically rules out a large pool of US postdocs, because life scientists there tend to earn their degrees much later than in Germany. And what about the candidate who took time out between graduating and doing a PhD to pursue other interests — perhaps an industrial stint? Or MD/PhDs, who are especially in demand for institutions wishing to enhance translational research? These individuals earn their degrees even later. Or what about European postdocs in the United States who would take a fellowship in Europe to get closer to home? One would think that a European postdoc who did a stint at, say, the National Institutes of Health would be quite desirable. But the age restriction would largely exclude such people.

One can try to understand the rationale for such restrictions. Perhaps the institutions consider scientists who get their PhD by a certain age more 'serious'? But by the same token, institutions that are serious about recruiting internationally need to lift such restrictions in order to get the quality of candidate they really want — regardless of age.

Paul Smaglik

Naturejobs editor



## Contents

### REGIONS

Texas goes for growth p1160

### CAREER VIEW

#### Scientists & Societies

Planning for a  
positive postdoc

#### Graduate Journal

Jason Underwood

#### Movers

Hidde Ploegh erererererp1162

### WWW.NATUREJOBS.COM

Career centre

Information on the  
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

IMAGE  
UNAVAILABLE  
FOR COPYRIGHT  
REASONS

# Thinking big Texas

J. SUGAR/CORBIS

Construction at the University of Texas Southwestern Medical Center continues unabated as it tries to keep pace with the demand for research space.

Texans will tell you that they have the biggest and best of anything. And increasingly, they're expanding their bragging rights from oil wells and cattle to biomedical research. The Lone Star State already has scientific size on its side. A study by the Institute for Economic Development at the University of Texas (UT) at San Antonio reported this March that the 15 members of the UT system contributed \$12.8 billion to the local economy in 2004. And the system's six health institutions alone provided 111,700 jobs.

But despite the size of its biomedical enterprise, Texas, which ranks third in the nation for university expenditure in the biosciences and is home to one of

the largest medical complexes in the world, has been slow to capitalize on underlying research strengths in biomedicine, biodefence and nanotechnology. That is changing as institutions look more squarely towards technology development, and actively promote research at the interface between the physical and life sciences.

One of the six medical-research institutions, UT Southwestern Medical Center in Dallas has grown from a small wartime medical college to an academic centre with 1,300 faculty members. The original campus couldn't contain this growth, so expansion has been focused on the newer north campus, which recently opened a 14-storey medical-research tower. Building work has started on a six-storey, 18,580-square-metre advanced-imaging centre for animal and human research in neuroscience and cancer. Plans and fund-raising are already under way for a seventh building on that site, says the university's president Kern Wildenthal.

This expansion is providing more scientific job opportunities, says Al Gilman, chairman of pharmacology and one of the medical centre's four Nobel laureates. Gilman expects the university to recruit about 130 research scientists to fill new buildings over the next five years. The medical-research tower will house up to 100 groups, including the entire departments of pharmacology and physiology, one-third of biochemistry, some neurology and developmental biology, and the Alliance for Cellular Signaling. And it will provide space for an expanded

UT SOUTHWESTERN MEDICAL CENTER



Harold C. Simmons Comprehensive Cancer Center.

Gilman also directs the Cecil H. and Ida Green Comprehensive Center for Molecular, Computational and Systems Biology, which will occupy part of the ninth floor. The centre, established in February 2004 with a \$12.8 million grant from the Cecil and Ida Green Foundation and Trust, will help UT Southwestern boost research in biological mathematical modelling. Gilman anticipates appointing half-a-dozen new staff members in systems biology.

With UT Southwestern's recent acquisition of nearby Zale Lipshy and St Paul university hospitals, the medical centre now owns and runs its hospitals. Administrators hope the move will place UT Southwestern among the ranks of the nation's top-tier academic medical centres. It should also broaden the scope of clinical trials and create opportunities to recruit clinical investigators.

In addition, UT Southwestern is beefing up research and clinical programmes in cancer, aiming for a Comprehensive Cancer Center designation from the National Cancer Institute. Its chances should be improved by the recent arrival of colorectal-cancer expert James Willson, who was recruited from Case Western Reserve University in Cleveland, Ohio. And plans are in the works to build a biotechnology incubator facility to nurture the development of start-up companies.



**Opportunities:** Al Gilman anticipates taking on more staff.

provide 20,700 square metres of new space, including an expanded stem-cell research programme. And, this month, the university broke ground on a new six-floor, 19,400-square-metre replacement research facility for the UT Medical School at Houston, to open in 2007.

Baylor College of Medicine, Houston's private research-intensive medical school, has 1,800 full-time faculty members. With research support of more than \$400 million, of which \$340 million is from federal sources, the college's strengths are in cardiology, cell and gene therapy, genomics and paediatrics. From July, its expanding neuroscience programme will be led by Michael Friedlander of the University of Alabama at Birmingham.

Exciting developments are at the interface between physics, engineering, maths and biology. Rice University, just west of Baylor, is noted for its research programmes in bioengineering and nanotechnology. Rice professors Richard Smalley and Robert Curl shared the 1996 Nobel Prize in Chemistry with Harry Kroto, of the University of Sussex, UK, for their 1985 discovery of fullerenes: a new form of carbon, C<sub>60</sub>.

Nanotechnology is founded on C<sub>60</sub> and carbon nanotubes (cylinders of carbon atoms measuring about a nanometre). More than 20 Rice researchers work in the field, especially at the Center for Nanoscale Science and Technology and the Center for Biological and Environmental Nanotechnology at Rice.

But Rice has no medical school, so Rice chemistry professor Bruce Weisman collaborates with researchers at UT Health Science Center at Houston, spending about a third of his time on biomedical applications of nanotubes. In 2002, he discovered that dozens of semiconducting nanotubes emit their own unique fluorescent signature. He wants to exploit this to develop non-invasive biological imaging agents.

Texas organizations such as the nascent Alliance for NanoHealth are also fostering collaborations between institutions in the medical centre. Going a step further, Rice plans to build a 46,500-square-metre collaborative research centre next to the medical centre.

Galveston, 80 km south of Houston, houses the University of Texas Medical Branch (UTMB). In the wake of the 2001 terrorist attacks and the anthrax letters that followed, UTMB formed a Center for Biodefense and Emerging Infectious Diseases to consolidate research in basic molecular and structural biology, tropical virology and vaccine development. In 2003, it was one of eight institutions chosen by the Department of Health and Human Services to become a Regional Center of Excellence for Biodefense and Emerging Infectious Diseases.

UTMB will also host a national biocontainment laboratory — one of only two large-scale biosafety level four labs to be located on a US university campus (the other will be in Boston). The \$150-million maximum containment facility will open in 2008.

The cost of living in Texas is quite low — and there is no state income tax. The biomedical enterprise in the Lone Star State seems to be thriving, so this is one case where a little Texan bragging may be justified. ■

**Diane Gershon** is assistant editor, technical reports, for *Nature Medicine*.

## BIOMEDICINE IN THE BAYOU CITY

But Dallas is not the only Texas city with ambitions of becoming a biotech hub. Houston boasts one of the largest medical complexes in the world. The 324-hectare Texas Medical Center houses more than 40 member institutions, including 13 hospitals and two medical schools, and in excess of 65,300 employees. The University of Texas's MD Anderson Cancer Center is now the top recipient of research grants from the National Cancer Institute — its federal research funds rose from \$63 million in 1998 to \$154 million last year.

The campus size has increased by about 50% over the past five years, and there are more buildings to come. Three new facilities opened in the past few months — an ambulatory clinical building, a cancer prevention building and the George and Cynthia Mitchell Basic Sciences Research Building.

At its 57-hectare south campus, the University of Texas is fostering development of a biomedical research park, spearheaded by MD Anderson president John Mendelsohn. The focal point will be Research Building 2, scheduled to open later this year. This will feature research programmes in gastrointestinal medical oncology and pathology, molecular pathology and molecular therapeutics. Building 1, open for a year, houses labs for immunology and cancer therapies.

Several building projects are also under way at the University of Texas Health Science Center at Houston. The future home of the Brown Foundation Institute of Molecular Medicine for the Prevention of Human Diseases, headed by Nobel laureate Ferid Murad, will

## Web links

UT Southwestern Medical Center at Dallas

♦ [www.utsouthwestern.edu](http://www.utsouthwestern.edu)

Texas Medical Center, Houston

♦ [www.tmc.edu](http://www.tmc.edu)

UT MD Anderson Cancer Center

♦ [www.mdanderson.org](http://www.mdanderson.org)

UT Health Science Center at Houston

♦ [www.uth.tmc.edu](http://www.uth.tmc.edu)

UT Austin

♦ [www.utexas.edu](http://www.utexas.edu)

UT Health Science Center at San Antonio

♦ [www.uthscsa.edu](http://www.uthscsa.edu)

UTMB

♦ [www.utmb.edu](http://www.utmb.edu)

UTMB's Center for Biodefense and Emerging Infectious Diseases

♦ [www.utmb.edu/CBEID](http://www.utmb.edu/CBEID)

Baylor College of Medicine

♦ [www.bcm.edu](http://www.bcm.edu)

Rice University

♦ [www.rice.edu](http://www.rice.edu)

Rice's Center for Biological and Environmental Nanotechnology

♦ [cben.rice.edu](http://cben.rice.edu)

Rice's Center for Nanoscale Science and Technology

♦ [cnst.rice.edu](http://cnst.rice.edu)

Alliance for NanoHealth

♦ [www.nanohealthalliance.org](http://www.nanohealthalliance.org)



## GRADUATE JOURNAL

## Leaving the family

My days in graduate school are numbered, and I'm not looking forward to the heartbreak of leaving my second family: my amazing lab group.

It's both a blessing and a curse that labs change people and flavours over the years. It's great to have new faces arrive with fresh perspectives and, especially in our lab, new recipes. Good skills at bench biochemistry often go hand-in-hand with good skills in the kitchen, and I've eaten some of the best food of my life in this lab.

It's hard to live in a dynamic workplace, though, when the family so often loses members as they move on. Sure, there will always be e-mails or phone calls, but the days of spontaneously running out for coffee, blasting gangster rap at 1 a.m. or playing a clever practical joke are over.

Now it is my turn to go and, although I won't miss the more tedious aspects of daily benchwork, I will miss the people who made the failures less painful and the successes more exciting. These people have been my cheerleading squad through thick and thin, in both my personal and professional life. Thanks for putting up with my neuroses, my dirty jokes and my pilfering of pens from your benches. The road to a PhD was long and a lot more fun with my second family. ■

**Jason Underwood is a graduate student in microbiology at the University of California, Los Angeles. He will graduate in June.**

## SCIENTISTS &amp; SOCIETIES

## Planning for a positive postdoc

If you want to give your postdoctoral position the best chance of being a success, you need a plan. That is one of the strong messages to come from a recent survey of US postdocs, conducted by Sigma Xi, the Scientific Research Society, in Research Triangle Park, North Carolina.

Over the course of a year, Sigma Xi assessed the productivity and workplace satisfaction of some 7,600 postdocs. Those who had sat down with their supervisors to draw up a plan at the start of their postdoc were more likely to have a happy and productive lab life, the survey found.

Specifically, the 72% of postdocs who made such plans were 40% less likely to be dissatisfied with their overall experience, 30% less likely to have had conflicts with their advisers, and

submitted 10% more papers for publication in peer-reviewed journals per year than those who did not.

The details of the plan also seemed to make a difference. The 39% of postdocs whose plan covered not only what they would do, but also what their advisers would do, were better off still.

Why do plans make such a difference? One explanation is that they are effective time-management tools that can help you to work more productively. A well thought-out research plan can focus your efforts and stop you from heading down blind alleys or working ineffectually.

A plan can also prevent disappointment and misunderstandings by setting your own and your adviser's expectations at an appropriate level from the outset.

An alternative explanation is that plans may be a good indicator of

the quality of a lab's management. Ask yourself who is more likely to be a good mentor: someone who sits down with you to draw up a career and research plan or someone who just turns you loose in the lab?

Indeed, 69% of those with a plan — and 80% of those whose plans included details of what their adviser would do — considered their advisers to be mentors, compared with only 48% of those with no plan.

So next time you are weighing up the pros and cons of a postdoctoral opportunity, you might do well to ask your prospective adviser about his or her management style. And you would certainly do well to check up on their track record for laying research and career plans for their postdocs. ■

**Geoff Davis is the principal investigator of Sigma Xi's postdoc survey.**

► [postdoc.sigmaxi.org/results](http://postdoc.sigmaxi.org/results)

## MOVERS

Hide Ploegh, Whitehead Institute for Biomedical Research, Cambridge, Massachusetts



Immunologist Hide Ploegh's career was helped along by a little bit of luck. As an undergraduate at the University of Groningen in the Netherlands, he was awarded a travel grant to work in Jack Strominger's lab at Harvard University. As luck would have it, this led him to the perfect place to uncover the molecular mechanisms of the immune system, and he has been

fascinated by the field ever since.

Ploegh's graduate years at Harvard set him firmly on his career path and exposed him to the latest technology. "People in our building were developing DNA sequencing techniques that we were able to take advantage of," he says.

Since then, he has been guided by his interest in biochemistry and immunology — seizing every available opportunity to pursue his research. This has allowed him to grapple with a wide range of issues from unravelling the intricacies of immune responses to foreign cells, to discovering how viruses manage to evade their host's defence mechanism.

Having returned to Europe, Ploegh's pivotal career move was probably his decision to leave a cancer institute in his native Netherlands and head back to the United States. He arrived at Massachusetts Institute of Technology (MIT), an experience he likens to being a

sous chef suddenly given a pantry full of the best ingredients and access to any possible technique.

After a few years at MIT, he began an eight-year stint at Harvard Medical School as head of its immunology programme. But now he is moving to the Whitehead Institute for Biomedical Research. With its links to MIT, this constitutes something of a homecoming for Ploegh — especially as it was the chance to exploit MIT's strength in materials science that lured him to the Whitehead. "In my research, there is an increasing importance for chemistry, materials science and microfabrication," he says.

As for careers advice, Ploegh maintains that the best route to success is simply to satisfy your curiosity. "I've come to realize that the best students are not necessarily those who have the most clearly laid career path, but those who dive in and do the work for the sheer joy of it," he says. ■

**CV** **1997–2005:** Professor of immunopathology and director of the graduate programme in immunology, Harvard Medical School, Boston, Massachusetts.  
**1992–97:** Professor of biology, Massachusetts Institute of Technology, Cambridge, Massachusetts.  
**1991:** Dean of graduate studies, Netherlands Cancer Institute, Amsterdam.  
**1986–92:** Head of department of cellular biochemistry, Netherlands Cancer Institute, Amsterdam.  
**1984–92:** Staff scientist, Netherlands Cancer Institute, Amsterdam.

# The Affinities

Be careful how you describe yourself.

**Robert Charles Wilson**

Of course I remember you. I want to clarify that point up front. Yes, Beth, of course I remember you — Montreal 2042 — and I'm flattered you think I can contribute to your project in even a modest way. As for your final query... well, we'll get to that. First let me address the professional questions.

*How did you discover the Affinities, and when did you join?*

I was at Columbia when Kleindeinst first achieved notoriety, and as an undergraduate psychology student I was aware of his work in epigenetic psychology and game theory. By that time he and his allies had already isolated some of the genes that we still call by their playful kleindeinstian names: *SKYDIVER*, *DARKROOM*, *SEXY* and so on; or *FEARFUL*, *COLLECTOR* and *GIMME* (all three of which had been correlated with obsessive-compulsive disorder). But I couldn't have guessed how influential his work would become, or the extremes to which it would be taken by earnest amateurs and clever entrepreneurs.

I became aware of the Affinities when the first purpose-built home sequencers came on the market. I knew K's work had progressed, that he had identified dozens of genes implicated in the development of personality. And I knew that he had sorted these into eight or ten clusters in which the genes interacted to suppress or promote one another — *DARKROOM* (associated with claustrophobia and agoraphobia) usually accompanied an active *COLLECTOR* (with its attention to detail and fear of disorder), for instance.

But how shocked poor Kleindeinst must have been when the first improvisational Affinity tribes, their names cobbled together from his whimsical genonyms, began to play the role astrology had once fulfilled for the general public. (*SEXY SKYDIVER FIREMAN* sounded better than Aquarius or Leo at the singles bar. And even if you considered yourself a *FEARFUL DARKROOM VOYEUR* you could seek out companionship in an Internet chatroom.)

The original Kleindeinst sequencer was introduced in 2038, with its venerable slogan "Know thyself", and the fad began in earnest. I bought mine later that year.

*Were the Affinity years a positive or negative experience for you?*

Much too big a question, Beth.

By 2040 I had established myself as a *GENTLE OPENMINDED COLLABORATOR*, the relevant subcodes (*VOYEUR* +, *VERBAL* + ...) enamelled on the Affinity badge I wore with no little pride. I chatted online with other *GOCs*, interacted preferentially with *GOCs* at work, and in autumn of that year I joined the local *GOC* community.

To encounter other souls so much like my own was like rediscovering a lost, beloved family. At Affinity gatherings there was much talk that

of recognition' even in a stranger's eyes.

To trust without fear. It was Edenic.

How could I not remember? You walked into the gilded lobby of the Hilton, your dark hair streaming with rain, and it was all inevitable from that point onward, as it had been inevitable a dozen times before, with dozens of other non-strangers. Our preferences were alike, our needs. Our deepest and most cherished desires. Ultimately, even, our mutual boredom.

*Should the Affinities be revived?*

Do you really need to ask?

They vanished for a reason, despite all the weeping. It wasn't just the conflicts predicted by game theory: the predatory Affinities like *DISHONEST MIMETIC PARASITE* that invaded *GOC* and stripped us of our fortunes and our easy mutual trust; it wasn't just the class-action suits accusing us of epigenetic favouritism. Those heartbreaks and betrayals were only the most public aspect of our terrible disappointment with, in the end, ourselves.

Lest nostalgia overwhelm us, recall that the last legal home sequencer was sold more than ten years ago; those that remain on the second-hand market are traded as quaint curiosities, like Victorian Ouija boards or phrenological models of the skull.

I wonder even at the wisdom of this research you're conducting. A book about the *GOC* Affinity? Really? Or just an excuse to renew old contacts?

*Gnothi seauton*, Beth. As the slogan says.

*Do you remember me, and shall we meet for coffee or drinks to discuss the book?*

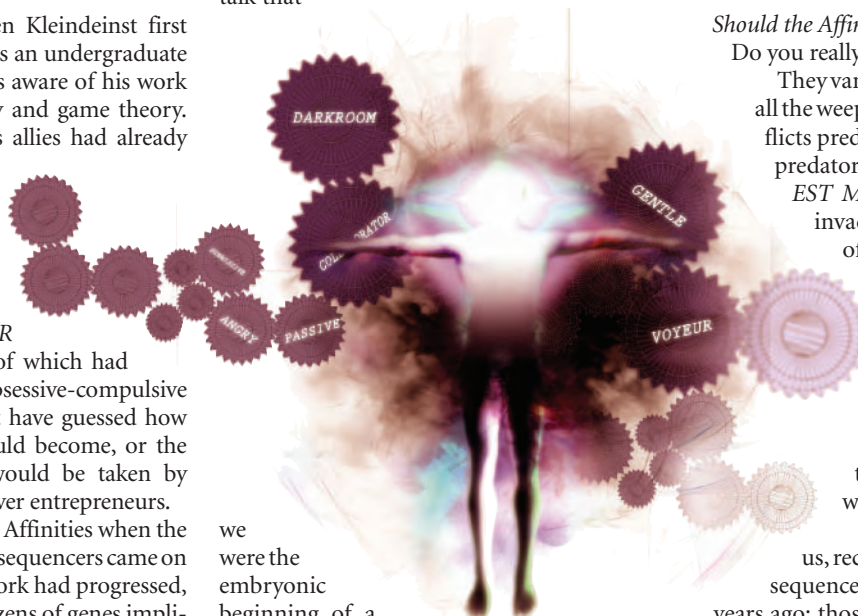
Of course I remember — but no, we won't meet.

I'm married now. My wife never belonged to an Affinity, even in the good years. She doesn't always laugh at my jokes, and our relationship is occasionally stormy.

But when I look into her eyes I do not see myself. And that dissonance, Beth — that impenetrable and always-surprising otherness — is, I have discovered, a blessing beyond price.

I'm sorry. But I can say without fear of contradiction that I know you'll understand. ■

Robert Charles Wilson has written several novels, including Hugo Award finalists *Darwinia* and *The Chronoliths*. His latest is *Spin*.



we were the embryonic beginning of a new speciation event — remember? The future, we thought, would be inhabited by wise *GOCs*, troll-like *AGGRESSIVE ANGRY DISSENTERS*, ethereal *PASSIVE SUBMISSIVE INNOCENTS*, and so on. What madness! And yet, how good it felt to flatter ourselves!

I booked a week off my medical practice to attend the 2042 *GOC* conclave in Montreal. We were in full flower then, weren't we, Beth? Because we were natural collaborators, we could interact without fear. We could make deals and hold ourselves to them without contracts; we could invest in businesses that weren't expected to be profitable for years, but which, when they did return a profit, made many of us independently wealthy. If we failed, we drew the appropriate lessons from our failures. We rarely argued, and we seldom even raised our voices.

And, need I say, we fell in love with one another. That was always the most delicious thing of all, wasn't it? To stroll down a hotel corridor full of people speaking dozens of languages and know for a fact that each one was a potential friend or lover: to see what literary scholars call 'the shock